(* COUNTS *)

29

One ordered response

(mixed two-way analyses)

		Exh	Page
Introduction	29-1	A	
		1	29-2a
29A.	Combining across cutting points 29-2	B	
	- concealment 29-4	2	29-5
		3	29-6
29B.	Parallel motion -- in pictures 29-4	4	29-8
	- a third example 29-10	5	29-9
		6	29-11
29C.	Combining as a two-way analysis 29-12	C	
	- demobilization plans again 29-16	7	29-13
		8	29-15
29D	Looking deeper 29-19	9	29-17/18
	- a three-way look 29-23	D	
		10	29-20
29E.	Quantitative responses are also ordered 29-26	11	29-22
	- an economic history example 29-28	12	29-24
		13	29-25
29F.	Running into an extreme 29-34	E	
	- a forest example 29-37	14	29-27
	- long scales 29-42	15	29-29/30
		16	29-31
29G.	Ordered circumstance and ordered response 29-47	17	29-33
	- marriage adjustment 29-47	F	
	- bootstrapping a scale 29-50	18	29-35
		19	29-36
	**********	20	29-39/41
		21	29-43
29H.	How far have we come? 29-56	22	29-44/45
		G	
	**********	23	29-48/49
		24	29-51
29K.	More pictures for the rain forest (optional) 29-58	25	29-53
		26	29-54
	**********	H	
		K	
29P.	Problems for chapter 29 29-63	27	29-59
		28	29-60
		29	29-61
		P	
		30	29-64
		31	29-65
		32	29-66
		33	29-67

29

One ordered response
(mixed two-way analyses)

Not all responses come as numbers. If the responses are those of a mixed fruit sorter: apple, orange, banana, pear, etc. we are not likely to do too well by ordering them a priori. There does not seem to be any natural order for such kinds of fruit.

Each time, however, there is a natural order among the possible responses, we can put it to good use.

Counted fractions -- so many did and so many didn't, so many succeeded and so many failed -- are often the simplest case of an ordered response. (There are exceptions -- the counting of Belgians as to whether French or Flemish is their native tongue is one such, since there can be no agreement as to which is higher or lower.) To have a counted fraction we need one cutting point. In many, many instances we have our choice of more than one cutting point, making the use of any single counted fraction a matter of choice. Indeed, so long as we know only how to use simple counted fractions, choosing among the alternative cutting points often becomes a major problem.

Learning how to use several cutting points, how to display information in terms of several cutting points, and

how to deal with situations just beyond the simplest kinds will be our aims in this chapter.

We will draw on almost everything we have learned about counted fractions -- all the techniques we met in the last chapter will have their analogs in this.

29A. Combining across cutting points

When we have several cutting points, we can of course use each one separately. Thus, for example, when soldiers about to be released from the army after World War II were asked about their plans to return to their previous employer (The American Soldier 1950, vol. 4, page 629), their answers were classified as: definite plans, tentative plans, considering returning, not considering returning. It would be reasonable to divide those more definitely returning from those less definitely returning at any one of three cutting points. Exhibit 1 gives the percentages corresponding to each of these three cuts under various circumstances. Each of these is a counted fractions, so that we naturally begin by converting them to f-roots. Once this is done in panel B of exhibit 1 we are ready to see whether the three cutting points lead us to similar conclusions.

Panel C of exhibit 1 shows the changes in f-root associated with each step of increase in duration of previous employment. In 5 out of 6 cases there is good agreement among the three cutting points. (The change from 1-2 years to 2-5

exhibit 1 of chapter 29 -- demobilization plans

% and f-roots for each of three cutting points as to plans to return to previous employer (see exhibit 2 for source)

Month of Separ.	Yrs. prev. empl.	Location of cutting point: Definite plans (α)	Tentative plans (β)	Considering return (γ)	Not Considering
A) % for STRONGER PLAN to RETURN -- stronger than cutting point					
July	<1	12	21	41	
	1-2	22	33	51	
	2-5	31	43	60	
	>5	49	65	79	
Sept	<1	18	40	57	
	1-2	35	59	78	
	2-5	47	64	77	
	>5	64	80	93	

Example: Of July separatees with < 1 year employment, 12% had definite plans, 21% had definite plans or tentative plans, and 41% had definite plans or tentative plans or were considering returning.

					(sums)
B) f-ROOTS for STRONGER PLAN to RETURN -- stronger than cutting point					
July	<1	-.84	-.61	-.18	(-1.63)
	1-2	-.59	-.35	.02	(-.92)
	2-5	-.39	-.14	.20	(-.33)
	>5	-.02	.30	.61	(.89)
Sept	<1	-.68	-.20	.14	(-.74)
	1-2	-.30	.18	.59	(.47)
	2-5	-.06	.28	.56	(.78)
	>5	.28	.63	.99	(1.90)

					(range)
C) CHANGE in f-ROOT					
July	<1 to 1-2	.25	.26	.20	(.06)
Sept	"	.38	.38	.35	(.03)
July	1-2 to 2-5	.20	.21	.18	(.03)
Sept	"	.24	.10	-.03	(.27)
July	2-5 to >5	.37	.44	.41	(.07)
Sept	"	.34	.35	.45	(.10)

years duration for September separatees shows consistent behavior.)

To the extent that all three cutting points show the same behavior, we are free:

- to use the results for any one of them. (In this example, if one cut were to be used, that between "plans" and "considering" would perhaps be most natural.)
- to combine them in what seems to be a useful way.

In this example, choosing different cut-offs gave us very similar final comparisons. To take full advantage of this fact, we need:

- to pull together the evidence from the various cuts into as good a summary as possible.
- to look at the discrepancies between individual results and summaries.

As always, we seek both summarization and examination of residuals. As always, we seek for residuals that are as close to individual values as we can make them.

Were we only interested in working summaries, trying to forget the various cutting points as soon as possible, it would be natural to replace each set of 3 values in panel B of exhibit 1 by:

- its mean, or
- its sum.

The sums are given in panel B; we shall return to them shortly.

* concealment *

We presented the data in exhibit 1 fairly, but, alas, not conventionally. The usual way to present such data is in terms of counted fractions for each individual level of response, just as though the responses were not ordered. Exhibit 2 gives this data, and data for a simpler example, as it is all too likely to be given.

When we see a two-way table of counts or counted fractions, we need to ask whether one or both factors really conceal response, and ought to be analyzed as such. In doing this, we will need to recognize that there are different kinds of responses, and thus different appropriate ways to analyze two-way tables in which concealed responses have been identified.

29B. Parallel motion -- in pictures

We seek to notice -- and then make use of -- similar behavior at different cutting points. The quality-of-production example is the simpler of those before us. Panels A and B of exhibit 3 give the cumulations and their f-roots. (We shall come to panel D in the next section.) What pictures should we look at in order to see the presence or absence of parallel motion among these f-roots?

Once we have ordered responses to which we do not want to give numbers, the problem of making pictures is not as it was in chapters 26 and 27. There we had a fixed scale

exhibit 2 of chapter 29 -- demobilization plans
quality of production

Examples of concealment of a response as a factor
in a two-way table

A) The DEMOBILIZATION-PLANS data as a TWO-WAY TABLE

Month of separation	Years previously employed	Return to previous employment: Definite plans	Tentative plans	Considering return	Not considering
July	<1	12%	9%	20%	59%
	1-2	22%	11%	18%	49%
	2-5	31%	12%	17%	40%
	>5	49%	16%	14%	43%
Sept	<1	18%	22%	17%	43%
	1-2	35%	24%	19%	22%
	2-5	47%	17%	13%	23%
	>>5	64%	16%	13%	7%

B) COUNTS -- number of instances of specified quality

Shift	Quality: a	b	c	d	e	f	(sum)
A	11	23	8	5	18	18	(83)
B	17	29	10	17	7	15	(95)
C	6	21	8	24	15	9	(83)
(sum)	(34)	(73)	(20)	(46)	(40)	(42)	((261))

C) SOURCES

(for panel A) The American Soldier 1950, vol. 4, page 629.

(for panel B) K. A. Brownlee 1949. Industrial Experimentation (4th ed.). Table on page 46.

exhibit 3 of chapter 29 -- quality of production

Brownlee's example of quality of product
(Presumably a is the best quality)

A) CUMULATIONS -- number of occurrences of at least specified quality, as given in exhibit 2.

	>a	>b	>c	>d	>e	(*)
A	11	34	42	47	65	(83)
B	17	46	56	73	80	(95)
C	6	27	35	59	74	(83)

Examples: 11 = 11, 34 = 11 + 23

B) f-ROOTS -- of corresponding counted fractions, as $\sqrt{f} - \sqrt{1-f}$

A	-.57	-.13	.01	.10	.34
B	-.48	-.02	.13	.40	.52
C	-.69	-.25	-.11	.31	.61

Example: $-.57 = \sqrt{11/83} - \sqrt{72/83}$

C) ELEMENTARY ANALYSIS -- two half-steps of median polishing, unit = 0.01 in raw f-root

						eff	fit
A	0	0	0	-21	-18	0	1
B	0	2	3	0	-9	9	10
C	-2	-2	0	10	19	-10	-9
eff	-58	-14	0	9	51	1	0
fit	-57	-13	1	10	52	0	-1

Hinges for residuals: -2 and +2

Clarity of fit: $\frac{128}{4} = 32.$

Outs: 5/15

D) PROBLEM

Repeat the calculations using f-logs.

and erected varying areas on given bases. Our problem is, instead, very much like that solved in section 28A, where we showed f-roots and f-logs with half cigars. There we had several cuts on a figure to illustate alternatives or compare before with after. Now we will have several cuts on a figure, one for each cutting point.

In this sort of picture, the area is fixed and the cuts move around, as shown in exhibit 4 for the extreme durations of various employment for July separatees from the demobilization plans data. (Recall that the table to make drawing the half-cigars easy is exhibit 5 of chapter 28.)

The difference between half-cigars for f-roots and for pluralities need not be great. That between half-cigars for f-logs and for pluralities is often great enough to make a very serious difference.

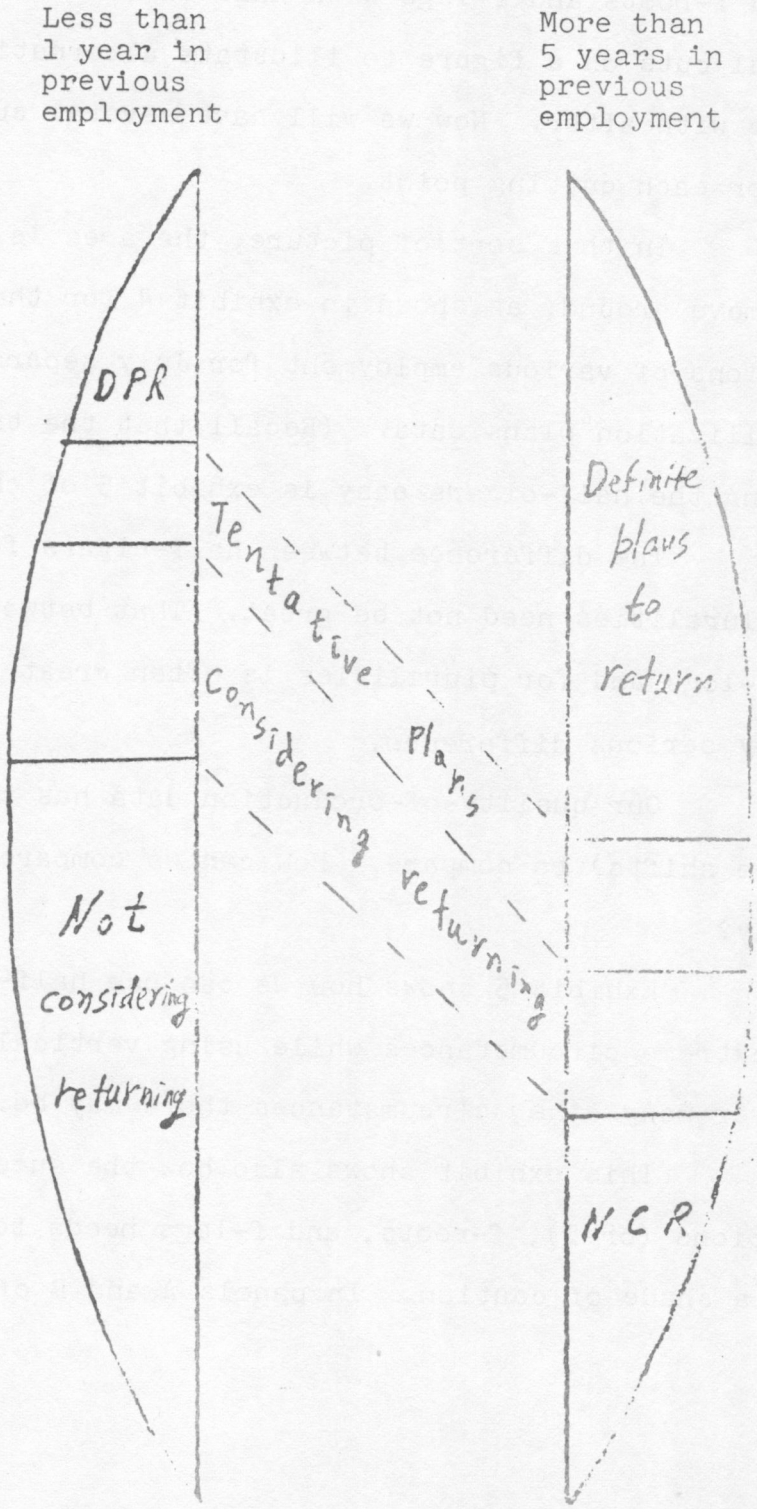
Our quality-of-production data has three circumstances (three shifts) to compare. How can we compare more than two easily?

Exhibit 5 shows how we can use half-cigars for the two extreme circumstances while using vertical lines for however many other circumstances there may be.

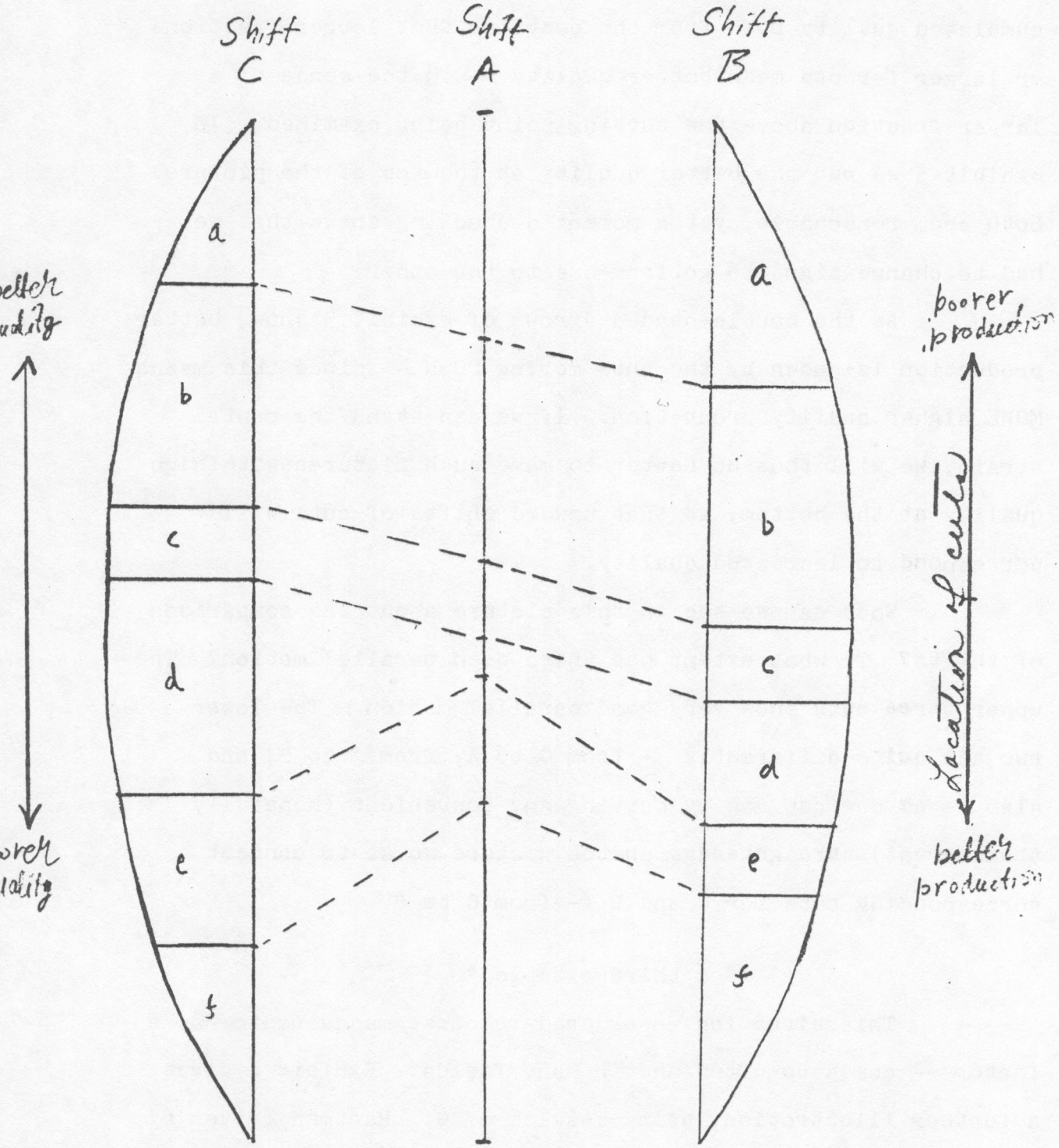
This exhibit shows also how the question of sign for fractions (or %), f-roots, and f-logs needs to be treated with just a shade of caution. In panels A and B of exhibit 3, we

exhibit 4 of chapter 29 -- demobilization plans

A graphical comparison of soldiers being demobilized in July 1945 with different amounts of previous employment (see exhibits 1 and 2)



Distribution of occurrences of specified quality
for each of the three shifts
(see exhibits 1 and 2)



PROBLEMS:

1) Copy this picture upside down.
2) Make a plot like this, probably upside down, for all 4 durations of previous employment for July separatees.
3) Do the same for September separatees.
4) Compare the results of (2) and (3) in words.

cumulated quality DOWN from the best, so that larger fractions or larger f-roots mean better quality -- in the sense of a larger fraction above the cutting point being examined. In exhibit 5 we put the better quality at the top of the picture. Both seem reasonable, yet a moment's checking shows that we had to change signs to go from one to the other.

As the double-headed arrows of exhibit 5 show, better production is shown by the cuts moving DOWN -- since this means MORE higher quality production. If we can stand the mental strain, we will thus do better to make such pictures with high quality at the bottom, so that upward shifts of cuts will correspond to increased quality.

What can we see in this picture about the comparison of shifts? To what extent has there been parallel motion? The upper three cuts show very good parallel motion. The lower two act quite differently -- from C to A, from A to B, and also -- as one can see by putting any convenient (hopefully transparent) straight-edge on the picture so as to connect corresponding cuts for C and B -- from C to B.

* a third example *

This situation -- ordered response masquerading as a factor -- comes up often and in many fields. Exhibit 6 gives a further illustration, using results of G. Hartman (given by Mather, 1947) concerning the tasting of phenylthiocarbamide by

exhibit 6 of chapter 29 -- sex in tasting

Hartman's data on testing phenylthiocarbamide
("Threshold" is the strongest solution not tasted)

A) The DATA

Threshold	Count: Men	Count: Women
10	15	42
9	35	52
8	46	38
7	31	30
6	23	19
5	13	17
4	9	6
3	7	5
2	10	10
1	13	19
0	25	33
<0	63	43
(Total)	(290)	(314)

B) SOURCE

K. A. Mather 1949. Statistical Analysis in Biology.

(Esp. Table 49 on page 196.)

C) PROBLEMS

1) Find counted fractions, f-roots, and differences of f-roots.

2) Make appropriate picture(s) of your results.

men and women. This is one of a number of substances for which the ability to taste varies widely and is inherited. Hartman used 11 strengths of solution, and called "threshold" the strongest solution that could not be tasted. (Thus <0 means that the weakest solution, 0, could be tasted.)

Conversion -- first to counted fractions and then to f-roots -- at each cutting point gives quite consistent results; the apparent difference between men and women varying irregularly between .21 and .10 (on the f-root scale) with men more sensitive. Clearly this analysis is doing well for us.

29C. Combining as a two-way analysis

In either the demobilization-plans example or the quality of production example we want to combine

in a resistant way

sets of three values that are offset one from another. If we are willing to be subjective, we can adjust the values rather arbitrarily, look at the results, set aside those seemingly discrepant, take means of the rest. Exhibit 7 does this for the demobilization-plans example in one way. Clearly we could do it somewhat differently -- and then argue as to which is best.

A moment's thought shows that what we are doing, namely

fitting

an adjustment for a cutting point

and

a value for conditions

is just a special case of two-way analysis.

exhibit 7 of chapter 29 -- demobilization plans

One form of summarization of the f-roots of exhibit 1

A) ADJUSTMENT and COMBINATION -- assumed adjustments in [ ]

Month	Duration	adjusted f-roots left-hand	middle	right-hand	Mean*
		[+.25]	[.00]	[-.35]	
July	<1	-.59	-.61	-.53	-.58
	1-2	-.34	-.35	-.33	-.34
	2-5	-.14	-.14	-.15	-.14
	>5	.23	.30	.26	+.26
Sept	<1	(-.43)	-.20	-.21	-.20
	1-2	-.05	.18	.24	.12
	2-5	.19	.28	.21	.22
	>5	.53	.63	.64	.60

(*) Mean of values not in parenthesis.

B) FITTED VALUES in 4×2 FORM

Duration	July	September	(Change)
<1	-.58	-.20	(+.38)
1-2	-.34	.12	(+.46)
2-5	-.14	.22	(+.36)
>5	.26	.60	(+.34)

We know how to make resistant two-way analyses. In fact, Panel C of exhibit 3 already did this for the quality-of-production data. Exhibit 8 shows the corresponding two-way plot, where -- since the vertical scale is in f-roots -- we can conveniently use a single half-cigar to aid visualization. This time we have put the poorer quality at the top, so that shift B's higher position indicates doing better.

Looking at the two-way analysis (panel C of exhibit 3) for full detail, we see that

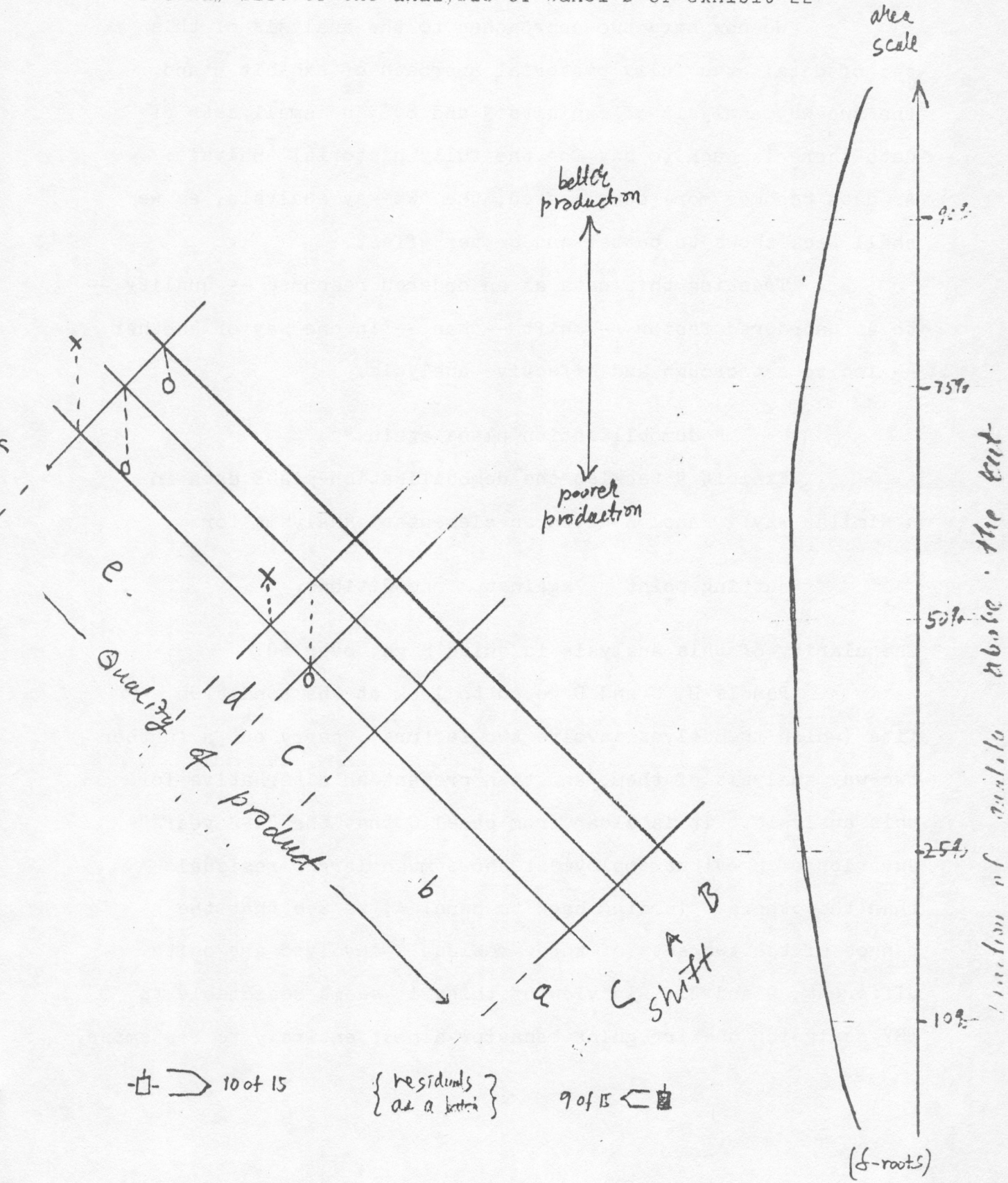
- shift B appears to perform somewhat better than shifts A and A better than C. (The small size of most residuals adds persuasiveness to this suggestion.)

- beyond this, shift C seems to have fewer occasions with very low quality -- qualities 'e' and 'f' together -- as shown by the residuals for the two lowest cuts: '≥e' and '≥d' -- while shift A seems to behave in the opposite way. (Might this be a peculiarity of the inspector or inspectors on these shifts?)

- the difference between cuts '≥a' and '≥b' and the difference between '≥d' and '≥e' seem to be greater than for other adjacent pairs of cutting points. Thus qualities 'b' and 'e' may extend over a wider range than qualities 'c' and 'd'.

exhibit 8 of chapter 29 -- shift production

Two-way plot of the analysis of panel D of exhibit 22



We now have two approaches to the analysis of this set of data: the fully pictorial approach of exhibit 5 and the two-way analysis of exhibits 3 and 8. For small sets of data there is much to say for the fully pictorial analysis. As data becomes more complicated, the two-way analysis, as we shall see, shows to better and better effect.

Treating this data as an ordered response -- quality -- to an unordered factor -- shift -- has -- in one way or another -- led to a thorough and effective analysis.

* demobilization plans again *

Exhibit 9 tackles the demobilization-plans data in a similar way. Panel A gives an elementary analysis for

cutting point against conditions.

The clarity of this analysis is quite high, over 60.

Panels B, C and D go on to look at the condition fits (which themselves involve two factors), carry out a further two-way analysis of them, and then present an alternative for this analysis. It is clear from panel C that the "1-2 years" duration of previous employment shows much larger residuals than the others. Turning back to panel A, we see that the ranges of the two sets of three residuals involved are quite different, 9 and 21. In view of this, it seems reasonable to TRY assigning the irregular behavior almost entirely to September.

exhibit 9 of chapter 29 -- demobilization plans

Condensing panel B of exhibit 1 by
resistant elementary analyses

A) Results of analysis by MEDIAN POLISHING -- unit = 0.01 in f-root

Month and duration		cutting point α	β	γ	eff	fit
July	<1	2	-8	0	-55	-53
	1-2	7	-2	0	-35	-33
	2-5	8	0	-1	-16	-14
	>5	1	0	-4	28	30
Sept	<1	-14	1	0	-23	-21
	1-2	-15	0	6	16	18
	2-5	0	1	-6	25	27
	>5	-2	0	1	61	634
	eff	-33	0	35	2	0
	fit	-31	2	37	0	-2

Hinges of residuals: -2 and H.

Clarity of fit: $\frac{184}{3} = 61.$

Outs: 7/24

B) CONDITIONS COMPARED -- fits from panel A

	July	Sept
<1	-53	-21
1-2	-33	18
2-5	-14	27
>5	30	63

C) ELEMENTARY ANALYSIS -- for panel B

	July	Sept	eff	fit
<1	2	-3	-36	-37
1-2	-7	7	-7	-8
2-5	-2	2	7	6
>5	2	-2	47	46
eff	-18	19	-1	0
fit	-19	18	0	1

Hinges for residuals: -2h and +2.

Clarity of fit: $\frac{121}{4h} = 27.$

Outs: 1/8

exhibit 9 (cont'd.)

D) A possible ALTERNATIVE

	July	Sept	eff	fit
<1	2	-3	-36	-37
1-2	-1	13	-13	-14
2-5	-2	2	7	6
>5	2	-2	47	46
eff	-18	19	-1	0
fit	-19	18	0	1

Hinges for residuals: -2h and +2.

Clarity of fit: $\frac{121}{5} = 27.$

Outs: 1/8

E) PROBLEM

1. Apply polishing by MEDIAN DIFFERENCES to the analysis of panel A. Show that the modified fits lead to hinges of residuals of -1 and +1h in the analog of panel C and thus to a clarity of 48.

This is done in panel D. As the problem indicates, a more refined analysis raises the clarity from 27 to 48. We shall return to this portion of the analysis in the next section.

Exhibit 10 portrays the two-way analysis of panel A of exhibit 9. Again we have used a portion of a split cigar to show -- by area -- how the fractions are indeed spread. As we, for instance, follow up the lowest NW-SE line of the two-way plot -- the one that corresponds to cut α -- we see the % with definite plans to return rising from about 10% (for July separatees with less than 1 year duration of previous employment) to about 65% for September separatees with more than 5 years duration). (Do the same for yourself along each of the other lines.)

The small size of the schematic plot of "residuals as a batch" in comparison with the very much larger two-way plot proper reminds us of the high clarity -- over 60 -- of this analysis. Clearly it is very reasonable to use the fit of this picture as a good summary of the data, even if 7 residuals out of 24 are outside.

29D. Looking deeper

As our analyses have hinted from time to time -- see exhibits 7 and 9 -- the circumstances in the demobilization-plans example form a two-way table. Month of separation and duration of employment are perfectly good factors. Indeed

exhibit 10 of chapter 29 -- demobilization plans

% stronger plans to return by cutting point, month of separation, and duration of previous employment
(fit of panel A of exhibit 9)

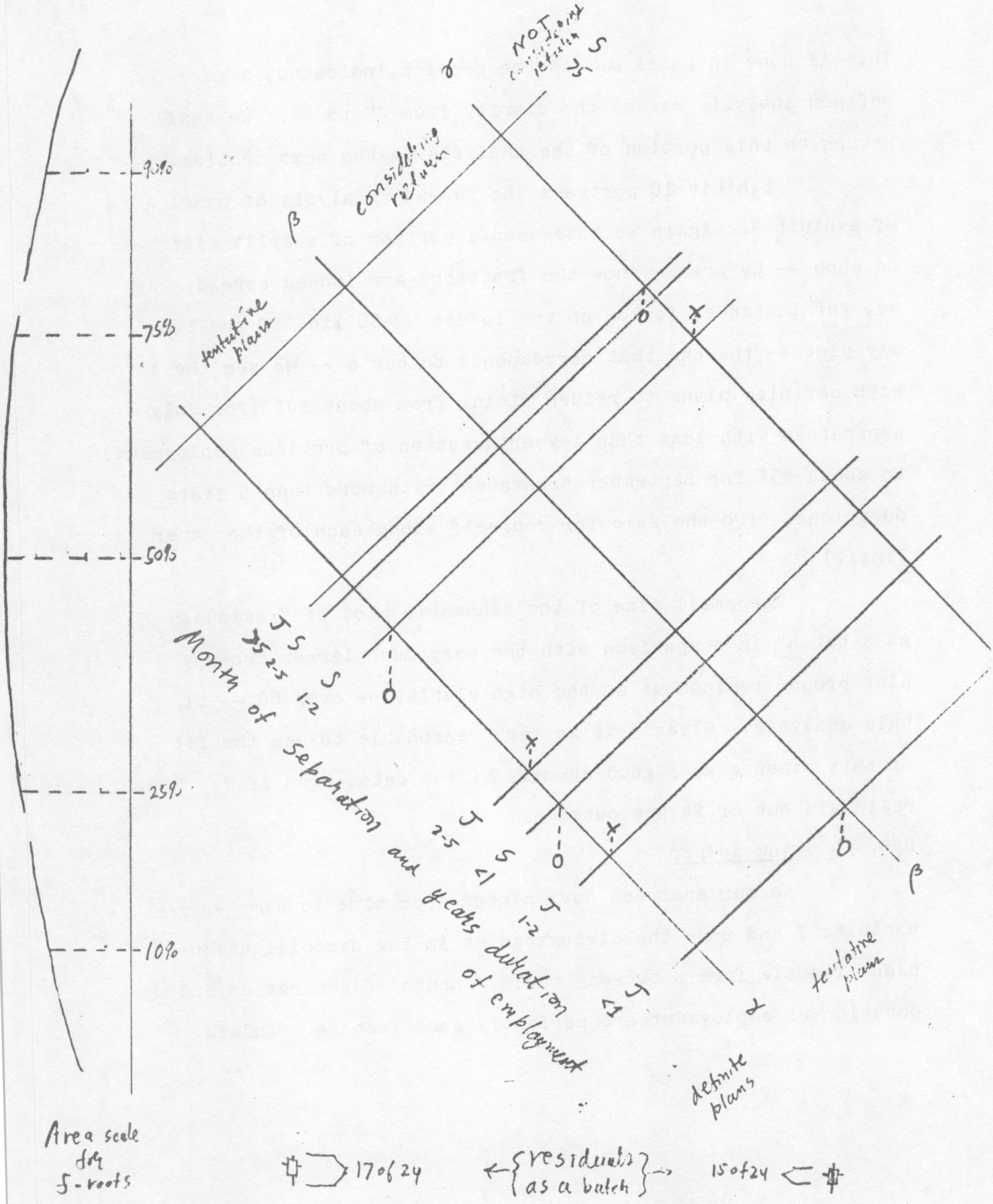


exhibit 9 has already laid out an elementary analysis for this two-way table, using as data the fits found by analyzing the response-circumstances table.

Exhibit 11 depicts the result. Since the effect for cut β is zero, we can take these overall fits as referring to cut β as we have done in exhibit 11, which is based on the [alternative] fit of panel D of exhibit 9 with its single outside residual. Again the picture tells us clearly that the fit is quite a good summary.

In this further analysis, based on the response-circumstance two-way analysis AS A FIRST STEP, we see one advantage of the two-way analysis of response and circumstance.

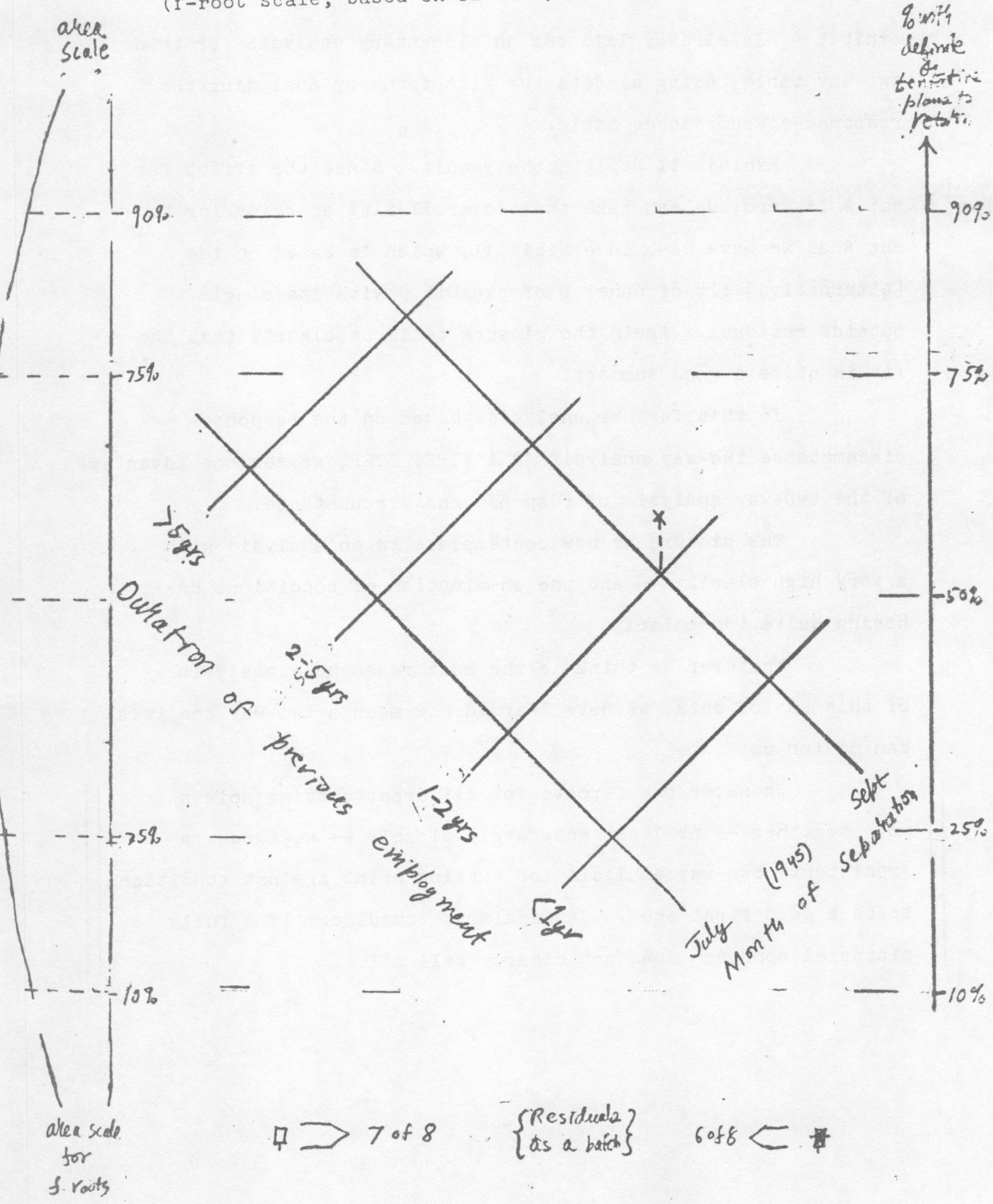
The picture we now contemplate is an analysis with a very high clarity -- and one combination of conditions behaving quite irregularly.

Whatever we think is the most reasonable analysis of this set of data, we have learned how much a two-way analysis can do for us.

Whenever the f-roots for different cutting points move together -- at least moderately closely -- we expect a (resistant) two-way analysis for cutting point against conditions to be a good first step. It is almost compulsory if a fully pictorial approach does not clearly tell all.

exhibit 11 of chapter 29 -- demobilization plans

Fitted % with tentative or definite plans to return
by duration of previous employment and month of separation
(f-root scale, based on fit of panel D of exhibit 9)



We ought also to notice how such a two-way analysis -- given reasonably parallel changes for the different cutting points -- essentially gives us a good start toward a scale for the response. Specifically, it gives us a scale for the cutting points. Panel A of exhibit 4, for example, states that the three cutting points we have used appear almost exactly equally separated, with separations of .33 and .35 in f-roots, respectively.

Exhibit 12 shows the simple picture of these separations. (Equal spacing does not always appear in such a situation.)

* a three-way look *

By now we have really described the original data as a three-way table where the pattern of factors is

response-by-circumstance-by-circumstance.

Exhibit 13 carries out the implied arithmetic. With only 7 constants fitted:

4 for duration fits,

1 for month effects (2 with median zero),

2 for cutting-point effects (3 with median zero),

this fit still has a clarity of 36, while the 3 (out of 24) outside residuals are clearly and markedly outside.

Since we have a single two-way fit to the two circumstance factors, to which IS ADDED an effect of cutting point, we can show the whole fit in a single picture. All we need do

exhibit 12 of chapter 29 -- demobilization plans

Scale for cutting points displayed vertically

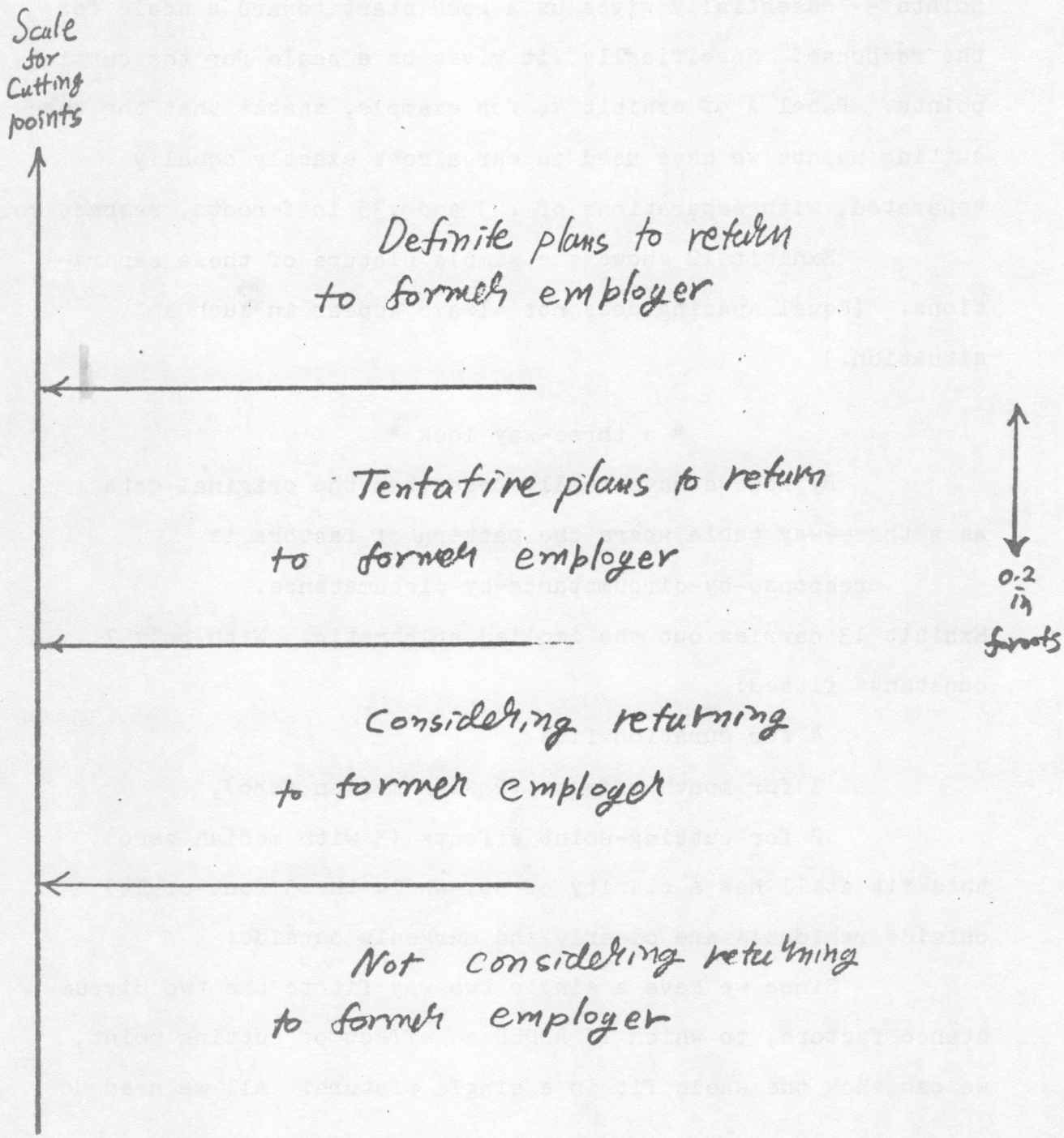


exhibit 13 of chapter 29 -- demobilization plans

Assembling the three-way fit, and finding
the associated residuals
(see exhibits 1, 2, 4, 7, 9 to 12)

A) ASSEMBLING the FIT for CIRCUMSTANCES -- entries in body of panel are sums of marginal entries, basis panel D of exhibit 9, unit -- 0.01 in f-roots

Duration of previous employment	(fit)	Month of separation	
		July	Sept
(eff)		(-19)	(18)
<1	(-37)	-56	-19
1-2	(-14)	-33	4
2-5	(6)	-13	24
>5	(46)	27	64

Examples: -56 = -19 + (-37), -33 = -19 + (-14), ...

B) ASSEMBLING the FIT for RESPONSE-by-CIRCUMSTANCES

	Month and duration	(fit)	Cutting point		
			α	β	γ
	(eff)		(-33)	(0)	(35)
July	<1	(-56)	-89	-56	-21
	1-2	(-33)	-66	-33	2
	2-5	(-13)	-46	-13	22
	>5	(27)	-6	27	62
Sept	<1	(-19)	-52	-19	16
	1-2	(4)	-29	4	39
	2-5	(24)	-9	24	59
	>5	(64)	31	64	99

Examples: -89 = (-33) + (-56), -66 = (-33) + (-33), ...

C) RESIDUALS -- observed f-roots are in panel B of exhibit 1, units still = 0.01 in f-root

July	<1	5	5	3
	1-2	-7	-2	0
	2-5	-7	-1	-2
	>5	-4	3	-1
Sept	<1	16	-1	-2
	1-2	1	14	20
	2-5	-3	4	-3
	>5	3	-1	0

Examples: Hinges of residuals = -2 and +3h.

Clarity of fit = $\frac{198}{5h}$ = 36.

Outs: 3/24

is to supply the two-way plot of the circumstance-by-circumstance fit with three different vertical scales, one for each cutting point. Exhibit 14 does just this, using half-cigar vertical scales for visual emphasis.

This picture shows us directly, and comparably, the effect of:

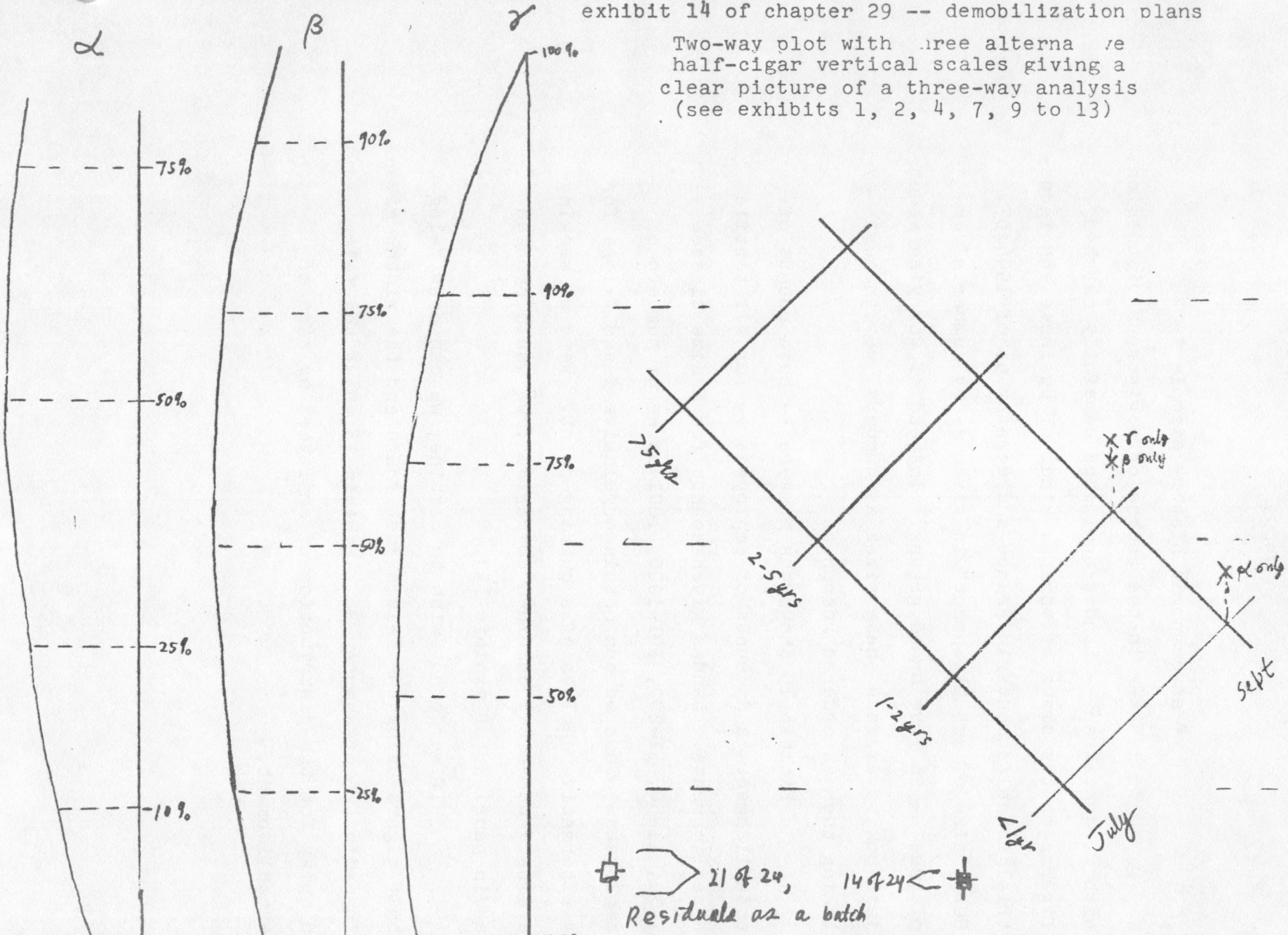
- changing the cutting point.
- changing the month of separation.
- changing the duration of previous employment.

29E. Quantitative responses are also ordered

When we have a response in numbers, it may pay us to begin the analysis by using less information than we seem to have. Especially if what we know about the responses is limited to knowledge of which of several broad classes each response falls in, we may do well by beginning our analyses as if all we knew about these broad groups were their order.

If an initial analysis of our responses -- actually available numerically in more or less broad classes -- as ordered responses is successful, we will acquire, as a by-product, a scale for the cutting points. We can then take this as a guide in re-expressing the original numbers for the cutting points. This can be both arithmetically simpler and less confusing than starting in by trying various re-expressions for the responses -- as given in broad classes of numbers.

Two-way plot with three alternative half-cigar vertical scales giving a clear picture of a three-way analysis (see exhibits 1, 2, 4, 7, 9 to 13)



α = definite plans
β = definite plans or tentative plans
γ = definite plans or tentative plans or

* an economic history example *

In 1954 -- in the course of a vigorous discussion with Gerschenkron -- Landes published carefully matched figures on the sizes of various kinds of business and industrial establishments in France and Germany as of 1906-07. Here "size of establishment" is given -- by number of employees -- in five broad groups. Instead of worrying about just how to express these sizes as numbers, we can begin by taking them as ordered responses.

Exhibit 15 shows the numbers -- both counts of establishments and counts of employees -- for all textile establishments. Landes gives counts for 5 size classes -- 1-10, 11-50, 51-200, 201-1000, 1001-up -- of numbers of employees. Thus we can obtain cumulative % and f-logs for 4 alternative choices of a cutting point. Were we seeking a single numerical combined comparison, we would proceed as in panel C (of exhibit 15).

If we do not wish to combine, we can draw f-log half-cigars using the f-logs just found and the widths given in exhibit 5 of chapter 28. Exhibit 16 shows the f-log display for the distribution of employees by size of establishment.

exhibit 15 of chapter 29 -- textile establishments

Sizes of textile establishment in
France and Germany, 1906-1907

A) ALL TEXTILE ESTABLISHMENTS

Count of employees	Count, etc., of establishments							
	France in 1906				Germany in 1907			
	(Count)	Count	%	f-log	(Count)	Count	%	f-log
(1001 up)	(48)				(71)			
1000.5		48	.0226	4.18		71	.0535	3.76
(201-1000)	(766)				(1013)			
200.5		814	.383	2.78		1084	.816	2.40
(51-200)	(1,676)				(2,748)			
50.5		2,490	1.17	2.22		3832	2.88	1.76
(11-50)	(3,636)				(4,984)			
10.5		6,126	2.88	1.76		8816	6.65	1.32
(1-10)	(207,501)				(123,768)			
(total)		(213,627)				(132,584)		

B) ALL TEXTILE WORKERS

Count of employees	Counts, etc., of individual workers							
	France in 1906				Germany in 1907			
	(Count)	Count	%	f-log	(Count)	Count	%	f-log
(1001 up)	(76,075)				(111,670)			
1000.5		76,075	8.27	1.20		111,670	10.20	1.09
(201-1000)	(296,026)				(392,756)			
200.5		372,101	40.43	.19		504,426	46.07	.08
(51-200)	(174,453)				(276,629)			
50.5		546,554	59.38	-.19		781,055	71.33	-.44
(11-50)	(88,469)				(119,326)			
10.5		635,023	68.99	-.39		900,381	82.23	-.69
(1-10)	(285,345)				(194,574)			
(total)		(920,368)				(1,094,955)		

exhibit 15 (Cont'd.) - 2

C) If we wished to COMBINE

Cut	f-logs for establishments France	Germany	Diff
1000.5	4.18	3.76	.42
200.5	2.78	2.40	.38
50.5	2.22	1.76	.46
10.5	1.76	1.32	.44
(median)			(.43)
	f-logs for individuals		
1000.5	1.20	1.09	.11
200.5	.19	.08	.11
50.5	-.19	-.44	.25
10.5	-.39	-.69	.30
(median)			(.18)

D) SOURCE

D. S. Landes 1954. Social attitudes, entrepreneurship, and economic development: A comment. Explorations in Entrepreneurial History 6:245-272. Appendix table 2 on pages 268-269.

exhibit 16 of cahpter 29 -- textile establishments

Split f-log comparison;
textile employees by size of establishment

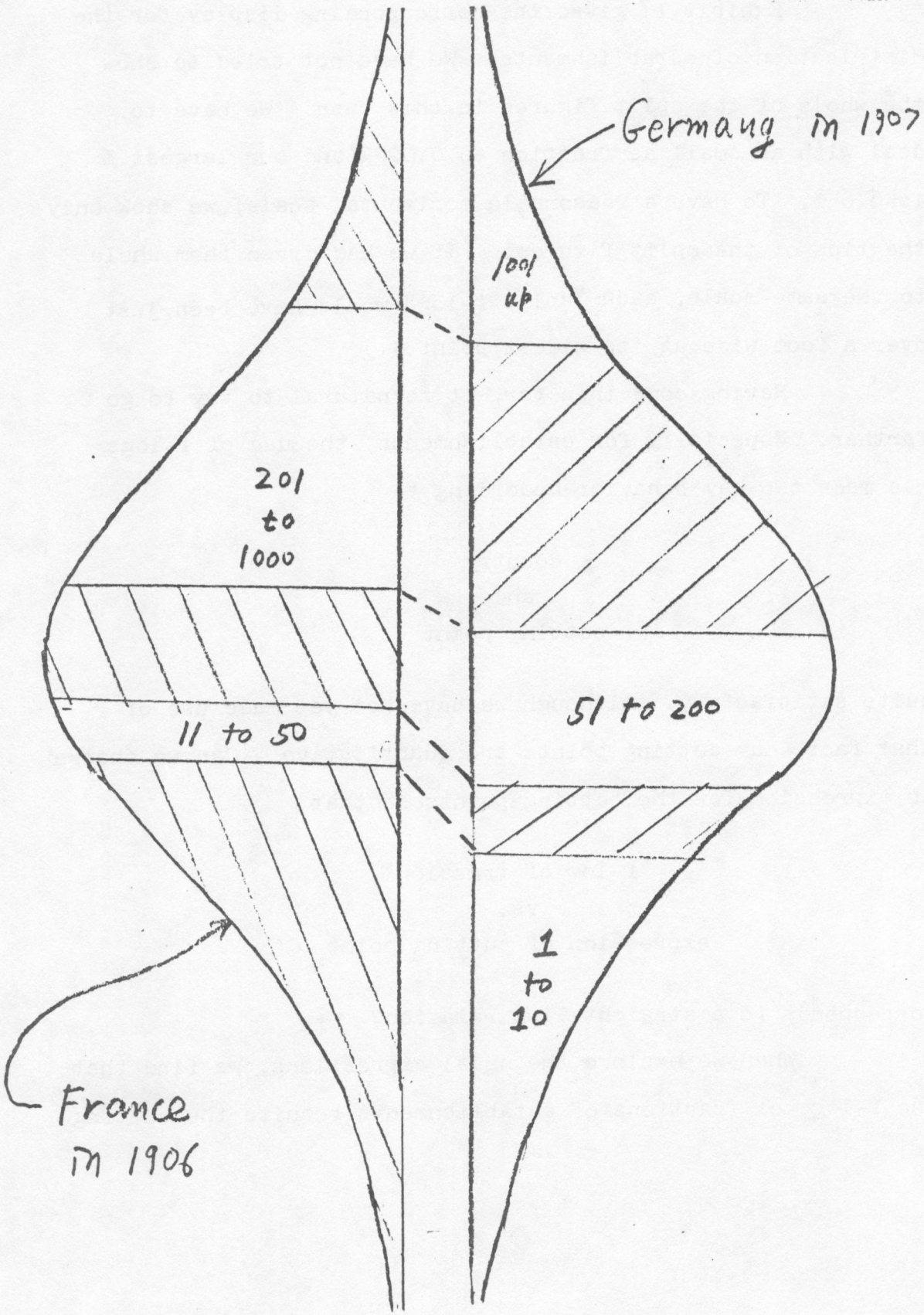


Exhibit 17 gives the corresponding display for the distribution of establishments. We have not tried to show the whole of the split figures in this case. We have to deal with as small as fraction as 0.022% and our largest % is 6.65%. To have a reasonable horizontal scale, we show only the tips of the split figures. (If we had drawn them whole to the same scale, each "split f-log" would have been just over a foot wide at its widest point.)

Having come this far, it is natural to try to go farther. Especially for establishments, the use of f-logs has made two-way behavior according to

country
and
cutting point

quite satisfactory. Although we have not yet made use of that fact, our cutting points are quantitative. Can we choose an expression for the cutting points so that

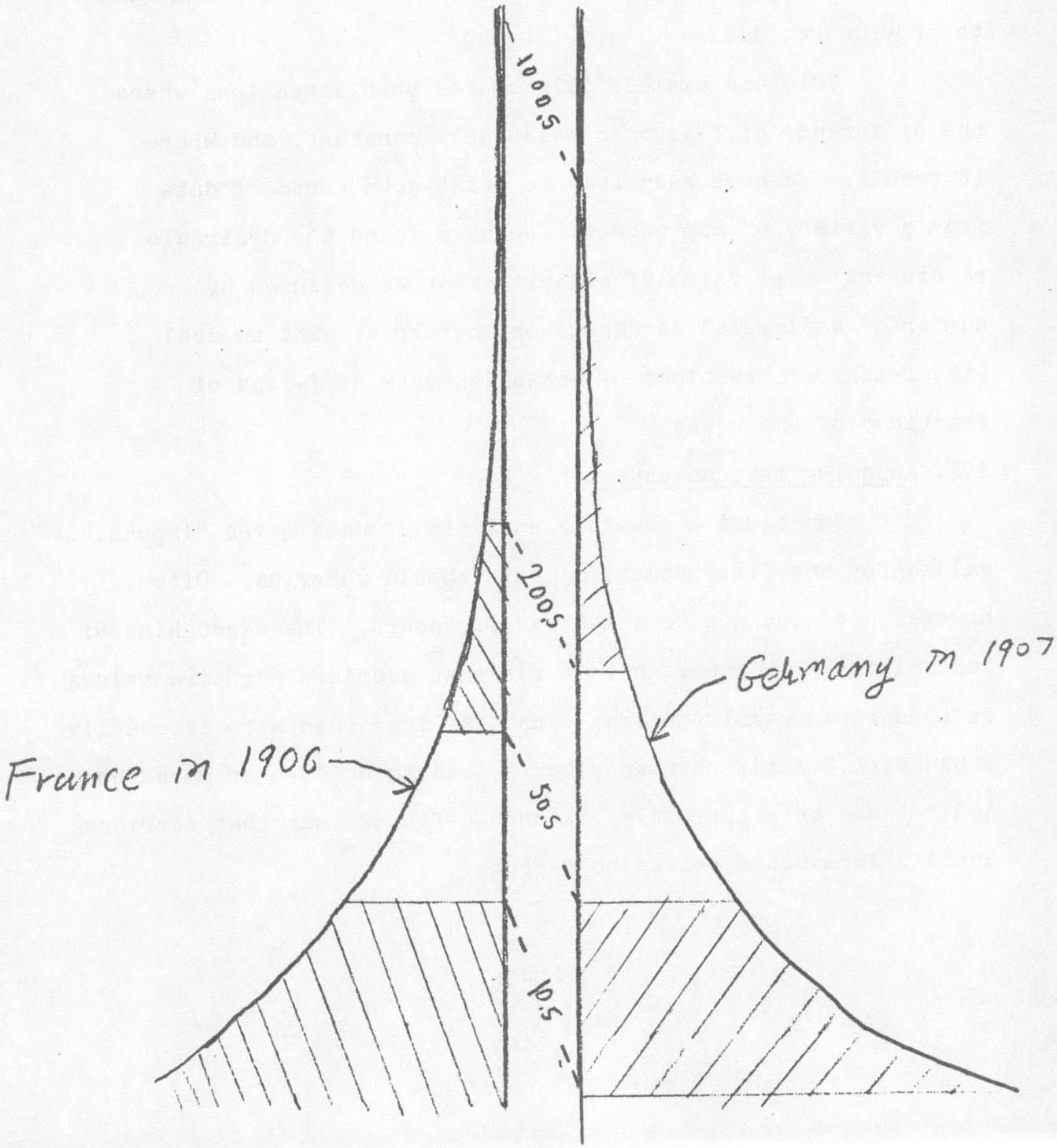
f-log of fraction
vs.
expression of cutting point

corresponds to a straight-line relation?

When we explore the usual expressions, we find that the f-logs of fractions of establishments require the cutting

exhibit 17 of chapter 29 -- textile establishments

Split f-log comparison of textile establishments by size



point to be expressed somewhere between logarithm and square-root. We naturally try the cube-root, with the results shown in exhibit 18. When we turn to the distribution of employees, we find that a square-root suffices, as shown by exhibit 19. While not as directly perceivable for many purposes as the half-cigar displays, these plots are much more ready and able to provide residuals.

This one example illustrates both situations where the difference of f-logs is reasonably constant, and where it trends. We have been able to treat both sorts of data from a variety of approaches. We have found the desirable re-expression of "size of establishment as measured by number of employees" to depend on whether we want to deal with f-logs of fractions of establishments of f-logs of fractions of employees.

29F. Running into an extreme

Sometimes a sensible analysis of data gives "impossible" values for the fit. Sometimes this should worry us. Often, however, it need not be a matter of concern. The exact kind of impossibility is important. A fit that ascribes negative values to a measured magnitude that cannot be less than zero is usually a bad sign. A fit that ascribes % less than 0% -- or more than 100% -- may be a perfectly reasonable fit, as may that ascribes equally impossible values to f-roots.

exhibit 18 of chapter 29 -- textile establishments

Another view of the size distribution of textile establishments

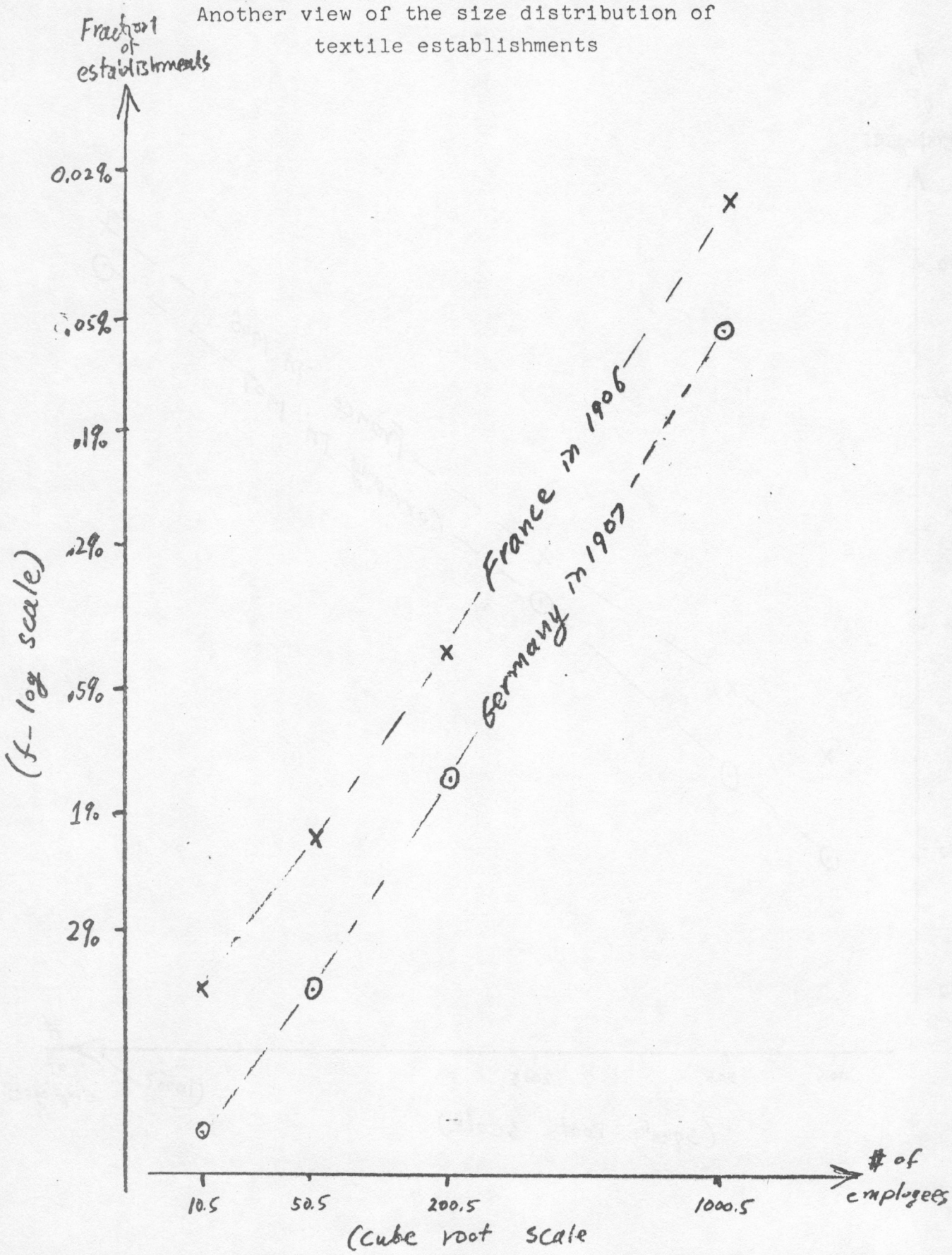
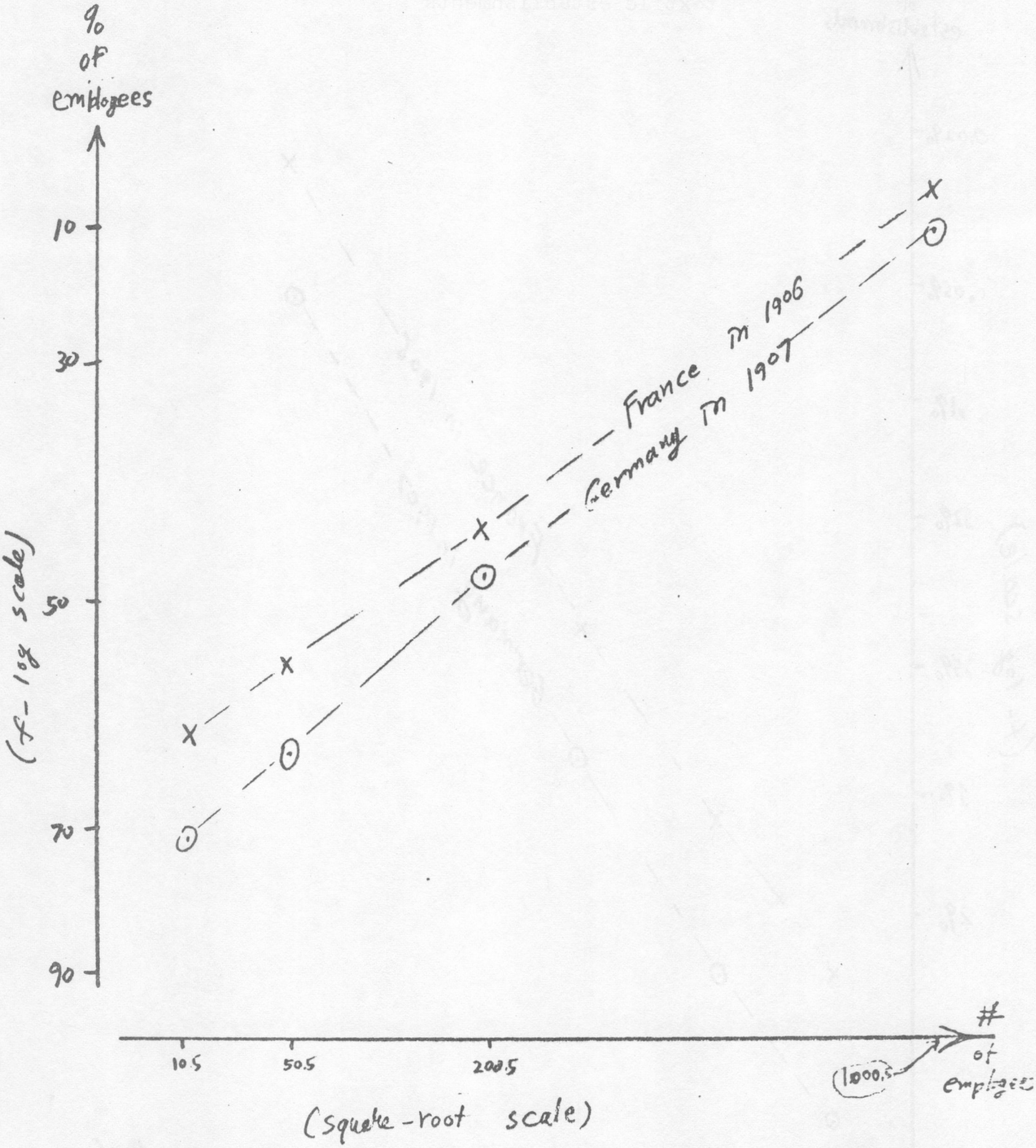


exhibit 19 of chapter 29 -- textile establishments

Another view of the distribution of employees
by size of establishment



The issue here is related to the famous statistical argument between G. Udny Yale and Karl Pearson as to whether "one man could be more dead of smallpox than another". (Try Karl Pearson, Biometrika, volumes 1 to 5, for references.) Which side do you think we are on?

A mosquito that was perched on the outside of the bomb casing in an atomic bomb test is "more dead", in a reasonable sense, than one swatted with a fly swatter. No small change would evade death in one case, two inches would in the other. "More than all" and "less than none" often have reasonable meanings.

* a forest example *

In following a developing piece of woods, it is natural to count the numbers of trees, not just by species, but by size. As a species enters it will be represented mainly by saplings -- then by small trees -- then by middle-sized trees -- finally by large trees. We measure size on a numerical scale, but we may do well to begin to analyze it as an ordered response. One natural way to follow progress in size is to divide all the trees of a given species among the size classes and calculate an f-root for each cutting point.

Next we wish to summarize these f-roots. This is often most simply done with a (resistant) two-way analysis either just resistant or resistant and modified.

Let us begin with data from one look at a forest where few changes are taking place. This will help us to become clearer about what we are doing, and what technical details need attention. Exhibit 20 has the counts of trees of different sizes for the eight commonest tree species in a single sample plot of Mixed Rain Forest at Moraballi Creek, British Guiana. Panels A to D take us to the f-roots, panel E to a two-way analysis, and panel F to a modified two-way analysis in which residuals coming from "all observed" -- an observed % of 100, observed f-root of 1.41 -- are set aside. Exhibit 21 pictures the result. (Note that we do not know how closely Davis and Richards measured their diameters; we do not know exactly how much above 4 inches the cut above the 0-4 inch class comes. Hence the label 4+ for this cut. Hence similar +'s on the others.)

Now we have an interesting complication -- one that we ought NOT let distress us. The fitted f-roots go above 1.41 thus seeming to call for more than 100% of that species of trees below that size. Clearly it is enough to call for ALL the trees to be no larger! So we may proceed easily with either of two attitudes:

exhibit 20 of chapter 29 -- rain forest

Diameter classes for the most common species
of trees in a mixed rain forest
Moraballi Creek, British Guiana

A) COUNTS -- by species and diameter

Species	counts by diameter in inches					
	0-4*	4-8	8-12	12-16	16-24	>24
Eschweilera (2 species**)	128	19	14	7	1	1
Eschweilera sagotiana	48	5	6	14	13	1
Licania laxiflora	24	18	13	5	6	-
Licania heteromorpha	176	29	12	-	-	-
Licania venosa	128	26	16	9	10	-
Ocotea rodiaei	16	8	2	5	10	3
Pentaclethra macroloba	80	31	20	16	6	-

* at least 15 feet high
**decolorans and pallida, not distinguished in the field.

B) CUMULATIONS -- up to given diameter

Species	≤4	≤8	≤12	≤16	≤24	(sum)
E2	128	147	161	168	169	170
Es	48	53	59	73	86	87
Ll	24	42	55	60	66	66
Lh	176	205	217	217	217	217
Lv	128	154	170	179	189	189
Or	16	24	26	31	41	44
Pm	80	111	131	147	153	153

C) PERCENTS -- up to given diameter

Species	≤4	≤8	≤12	≤16	≤24
E2	75.4	86.5	94.5	98.8	99.4
Es	55.1	61.0	68.9	83.9	98.8
Ll	36.5	63.7	83.3	90.9	100.0
Lh	81.1	94.5	100.0	100.0	100.0
Lv	67.6	81.5	90.0	94.7	100.0
Or	36.4	54.5	59.0	70.5	93.2
Pm	52.6	73.0	86.0	96.5	100.0

(slide-rule accuracy)

(cont'd.)

exhibit 20 cont'd.

D) f-ROOTS -- for percents

Species	≤4	≤8	≤12	≤16	≤24
E2	.53	.80	1.04	1.25	1.30
Es	.10	.22	.39	.73	1.25
Ll	-.27	.28	..71	.92	1.41
Lh	.66	1.04	1.41	1.41	1.41
Lv	.36	.67	.89	1.05	1.41
Or	-.27	.09	.18	.42	.99
Pm	.05	.47	.79	1.12	1.41

E) TWO-WAY ANALYSIS -- 4 half-steps of median polishing, unit 0.01 in matching f-roots.

Species						eff	fit
E2	5	0	0	0	-44	33	104
Es	14	-6	-13	0	3	-19	52
Ll	-42	-19	0	0	0*	0	71
Lh	0	6	19*	-2*	-51*	51	122
Lv	3	2	0	-5	-18*	18	89
Or	0	4	-11	-8	0	-42	29
Pm	-10	0	8	20	0*	0	71
eff	-56	-24	0	21	70	71	00
fit	15	47	71	92	141	00	-71

Hinges of residuals: -8 and 3. Clarity of fit: $\frac{249}{11} = 20$.
Outs = 4/36.

* indicates residual coming from an observed f-root of 1.41

If we omit the *ed residuals, we find

Hinges of residuals: -7 and 3. Clarity of fit: $\frac{201}{10} = 20$
Outs = 5/29.

If we squash fits >1.41 down to 1.41, the first clarity becomes $\frac{168}{11} = 15$.

(cont'd.)

exhibit 20 cont'd.

F) MODIFIED TWO-WAY ANALYSIS -- *ed values omitted from medians, 1 half-step of median polishing**

						eff	fit
E2	5	0	0	0	-44	33	104
Es	14	-6	-13	0	3	-19	52
Ll	-32	-9	10	10	10*	-10	61
Lh	-3	3	16*	-5*	-53*	54	125
Lv	2	1	-1	-6	-19*	19	90
Or	0	4	-11	-8	0	-42	29
Pm	-10	0	8	20	0*	0	71
eff	-56	-24	0	21	70	71	00
fit	15	47	71	92	141	00	-71

Hinges of residuals:*** -7 and +3h.

Clarity of fit*** = $\frac{174 - (-33)}{10h}$ = 20.

Outs*** = 3/29.

* indicates residual coming from an observed 1.41.
** indicates no attempt to force median of species effects to zero.
***indicates omitting all *'d residuals.

If we squash all fits above 1.41 down to 1.41, the clarity*** becomes $\frac{168}{10h}$ = 16.

G) SOURCE

P. W. Richards 1952. The tropical rain forest. An ecological study. Cambridge University Press. Table 7 at page 53.

His source:
T. A. W. Davis and P. W. Richards 1933-4. The vegetation of Moraballi Creek, British Guiana; an ecological study of a limited area of Tropical Rain Forest. Parts I and II. Journal of Ecology 21: 350-384 and 22: 106-155.

- a matched f-root of more than 1.41 calls for no more than a matched f-root of 1.41, namely "all there are"; a matched f-root of less than -1.41 calls for "none".

- our fit is to be defined as "the sum of row effect, column effect, and common term, reduced to 1.41 if this sum is above 1.41 and increased to -1.41 if the sum falls below.

Whichever choice we take, we can proceed blithely with our analyses. (We may have to be careful to warn the computer if we are using one.)

We remarked that any such analysis gives a scale for the cutting points between our ordered classes. Exhibit 21 plots the fits for the cutting points against their size in inches. The relation is a nice straight line. For this data, at least, diameter in inches is a good way to measure size. (This example will be taken further in optional section 29K.)

* long scales *

We can still use similar techniques when the data is much more sparse. Our two-way analysis may, as in exhibit 22, be forced to concentrate on differences between f-roots or f-logs for adjacent cuts. In this rather extreme made-up example, for instance, there is no place where we can compare nonadjacent cuts.

exhibit 21 of chapter 29 -- rain forest

Fit for cutting point plotted against the corresponding diameter in inches
(see exhibit 20)

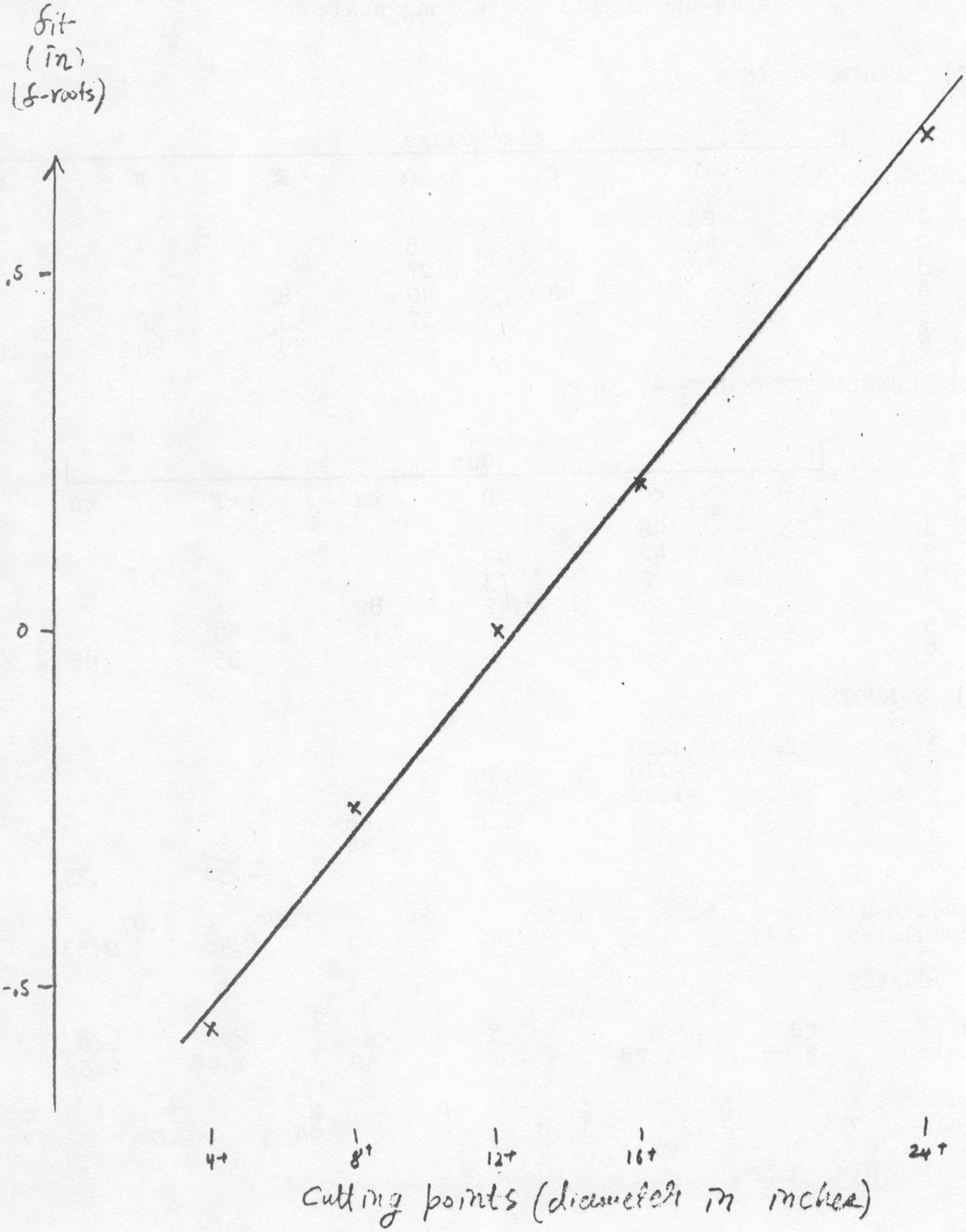


exhibit 22 of chapter 29 -- casual example

Made-up example for a long scale

A) COUNTS -- in %

Area	Size class A	B	C	D	E	F	G
1	75	23	2				
2		42	50	8			
3		2	71	27			
4			42	40	18		
5				12	50	38	
6					35	50	15

B) CUMULATIVE COUNTS -- in %

	Cut <B	<C	<D	<E	<F	<G
1	75	98				
2		42	92			
3		2	73			
4			42	82		
5				12	60	
6					35	85

C) f-ROOTS

1	.52	1.20				
2		-.16	.96			
3		-1.20	.47			
4			-.16	.68		
5				-.84	.30	
6					-.30	.77
(median diff)	.68	1.36	.84	1.04	1.07	
cumulated	-2.46	-1.78	-.42	.42	1.46	2.53

D) SCALES

cuts	<B	<C	<D	<E	<F	<G
scale	-2.46	-1.78	-.42	.42	1.46	2.53

centers	A	B	C	D	E	F	G
scale	??	-2.12	-1.10	0	.94	2.00	??

Trial scale

-3	-2	-1	0	1	2	3

exhibit 22 - cont'd.

E) SOURCE

Invention.

F) PROBLEM

Make up a set of counts in the pattern of panel A that will not lead to equal spacing and carry through the calculations.

In this example, we have little evidence for parallel motion at different cut points. The only place we can compare differences is between <C and <D where areas 2 and 3 give 1.06 and 1.67, respectively. Not a very close check.

We can, however, still paste together the few differences between size classes we are given to reach the cumulated values at the close of panel C. In panel D these are taken as scale values for the cuts, which are then converted into scale values for centers. In this made-up case, these center values are so nearly equally spaced as to make it most natural to use equally spaced scale values for classes A to G. As always, we have little evidence where the centers of the end classes fall, but we do our best with what we have.

With other sets of data, the resulting scale will not proceed by equal steps, as the reader should check.

Such scales are likely to be not only long but loose and floppy. Too often, though, they are all we have. When in doubt, use the best -- the best you have!

29G. Ordered circumstance and ordered response

Having one factor in a two-way table an ordered response does not keep the other factor from being ordered. When this other factor does not conceal a response, we begin as before, reducing the ordered response to a set of f-logs (or f-roots), and summarizing each set. We then take advantage of the ordered nature of the other factor in analyzing the summaries.

* marriage adjustment *

In 1936, Burgess and Cottrell gave the data on the relation between husband's education and marriage-adjustment scores shown in exhibit 23. Clearly we are allowed to take marriage-adjustment score as a response.

Whether this score is a response to education itself, to that increased ability which usually accompanies greater education (perhaps as a cause, not an effect), or to the common effects of parental attitudes on desire for education and ability to adjust, is not a question to be answered from this set of data. One reader suggested that it might be the effect of well-adjusted wives putting husbands through college. (If we wished to treat both education and adjustment as responses, we would need techniques such as those we will meet in the next section.)

exhibit 23 of chapter 29 -- marriage adjustment

Counts of combinations of husband's education and marriage adjustment scores

A) COUNTS

Husband's Education	Marriage adjustment score: Very low	Low	High	Very high	(Sum)
Graduate Work	4	9	38	54	(105)
College	20	31	55	99	(205)
High School	23	37	41	51	(152)
Grade School	11	10	11	19	(51)

B) CUMULATED COUNTS

	≤ Very low	≤ Low	≤ High	(Sum)
Graduate Work	4	13	51	(105)
College	20	51	100	(205)
High School	23	60	101	(152)
Grade School	11	21	32	(51)

Examples: 4 = 4, 4 + 9 = 13.

C) f-ROOTS -- as $\sqrt{f} - \sqrt{1-f}$

Graduate Work	-.78	-.58	-.02
College	-.64	-.37	.02
High School	-.53	-.15	.24
Grade School	-.42	-.12	.18

Example: $-.78 = \sqrt{4/105} - \sqrt{101/105}$.

D) ELEMENTARY ANALYSIS -- of panel C, by median polish, unit = 0.01 in f-roots

				eff	fit
Graduate Work	0	-11	6	-21	-47
College	4	0	0	-11	-37
High School	-	0	0	11	-15
Grade School		0	-9	14	-12
eff	-31	00	39	-26	00
fit	-57	-28	13	00	26

Hinges for residuals are -3h and +0h. Clarity of fit = $\frac{105}{04}$ = 26.

exhibit 23 (Cont'd.) - 2

E) SOURCE

E. W. Burgess and L. S. Cottrell 1936. The prediction of adjustment in marriage. Amer. Sociol. Rev. 1:737-751.

So again we ought to accumulate counts, convert to f-roots or f-logs and proceed to an elementary analysis. Except for a few isolated irregularities (-.11 among Graduate Work, -.09 among Grade School) that appear to be concentrated in the smallest classes, the fit by stripes seems quite good. The most meaningful results are the stripe adjustments for education, namely

Graduate Work:	-.21
College:	-.11
High School:	+.11
Grade School:	+.14

which show a steady shift (in the direction of better adjustment) with increasing education. Exhibit 24 depicts our results.

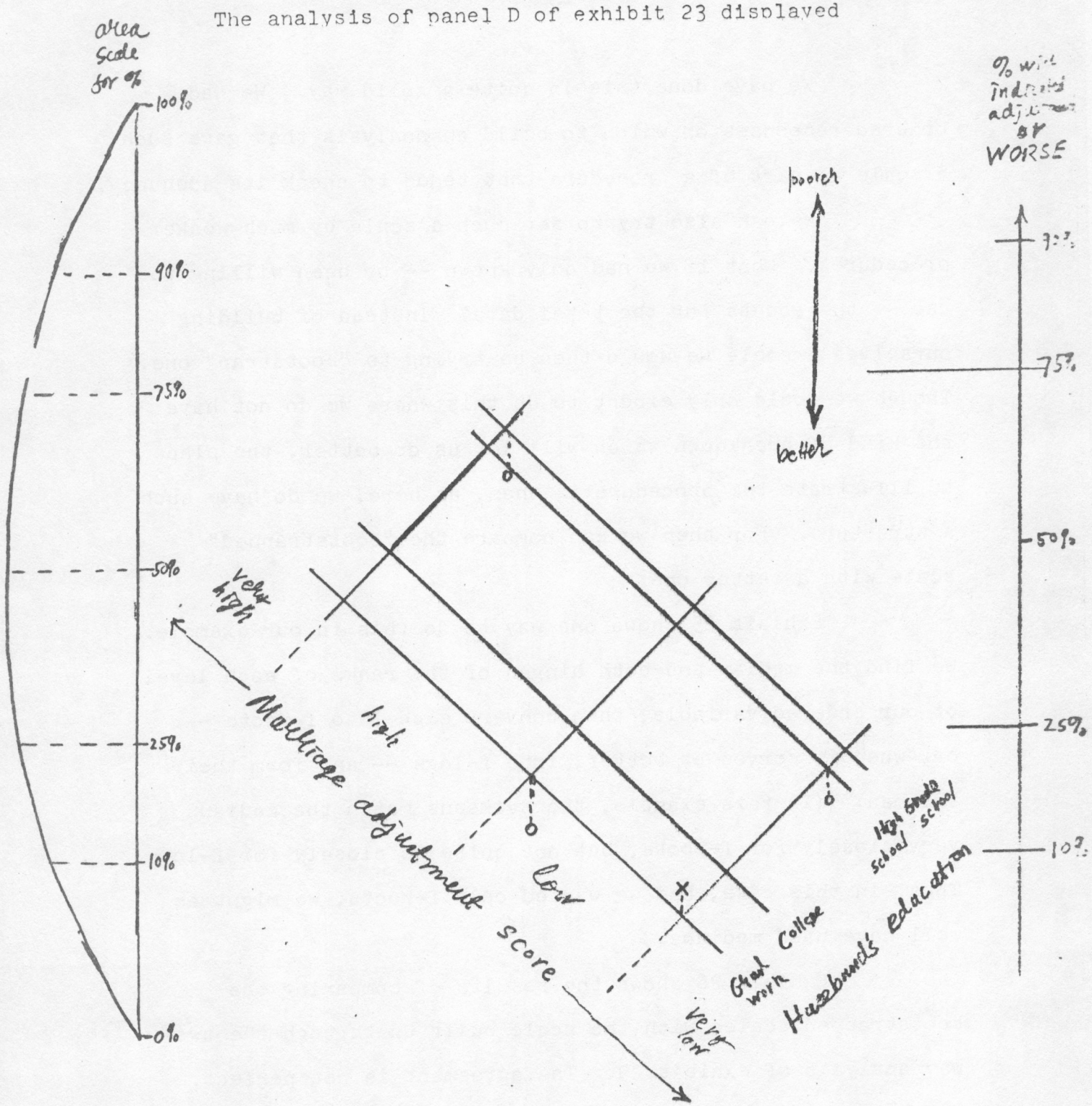
* bootstrapping a scale *

So far so good, but we will clearly be better off -- for two reasons -- if we can replace the ordered variable "husband's education" by a wisely numbered variable, namely:

- in situations where either the effect is small or many different levels of the ordered factor occur, we can make a more sensitive inquiry if we can look explicitly for tilt, rather than just looking for an ordered trend.

- in situations -- like our example -- where there is no difficulty spotting a trend, we have a better chance to ask about its character if we can plot numbers against numbers.

The analysis of panel D of exhibit 23 displayed



8 of 12

residuals as a batch

6 of 12

We have done this in quite a solid way. We had ordered responses on which to build an analysis that gave such a scale as part of a procedure that tends to check its adequacy.

We can also try to set such a scale by much weaker procedures. What if we had only known -- or been willing to use -- the counts for the total data? Instead of building ourselves a scale we would then be trying to "bootstrap" one. Though we would only expect to do this where we do not have the kind of breakdown which will let us do better, the place to illustrate the procedure is when, as here, we do have such a structure. For then we can compare the "bootstrapped" scale with a better one.

Exhibit 25 shows one way to do this in our example. We find the median and both hinges of the ranks of each level of our ordered variable, then convert each into f-roots -- or, when it serves us better, into f-logs -- and form their trimean. (In this example, the trimeans match the medians very closely for f-roots, but not quite so closely for f-logs. Thus, in this case, had we wanted only f-roots, we might as well have used medians.)

Exhibit 26 shows the results of comparing the bootstrapped scales with the scale built up through the two-way analysis of exhibit 25. The agreement is not perfect. We could not expect it to be. It is close enough, however, to strengthen our belief in both scales.

exhibit 25 of chapter 29 -- marriage adjustment

"Bootstrapping" a scale from totals only.
Calculation of plotting values for four levels of husband's education

Husband's education				f-roots		f-logs	
In words	In inverse ranks: Covered	HMH		HMH	trimean	HMH	trimean
Graduate Work	1 to 105		27	-1.05		-1.44	
			53	-.88	-.88	-1.08	-1.10
			79	-.72		-.81	
College	106 to 310		157	-.40		-.41	
			208	-.19	-.20	-.19	-.20
			259	.01		.01	
High School	311 to 462		348.5	.37		.38	
			386	.52	.53	.56	.57
			424.5	.70		.78	
Grade School	463 to 513		475	.97		1.27	
			488	1.06	1.07	1.48	1.52
			501	1.18		1.86	
(Total)			(513)				

HMH = hinge, median, hinge.

Examples: 105 = 105, 310 = 105 + 205

$53 = \frac{1}{2}(1+105)$, $27 = \frac{1}{2}(1+53)$, $79 = \frac{1}{2}(53+105)$

-1.05 and -1.44 from 5.25% = 27/513 through exhibit 1 of chapter 28

$-.88 = \frac{1}{4}(-1.05 + 2(-.88) + (-.72))$

$-1.10 = \frac{1}{4}(-1.44 + 2(-1.08) + (-.81))$

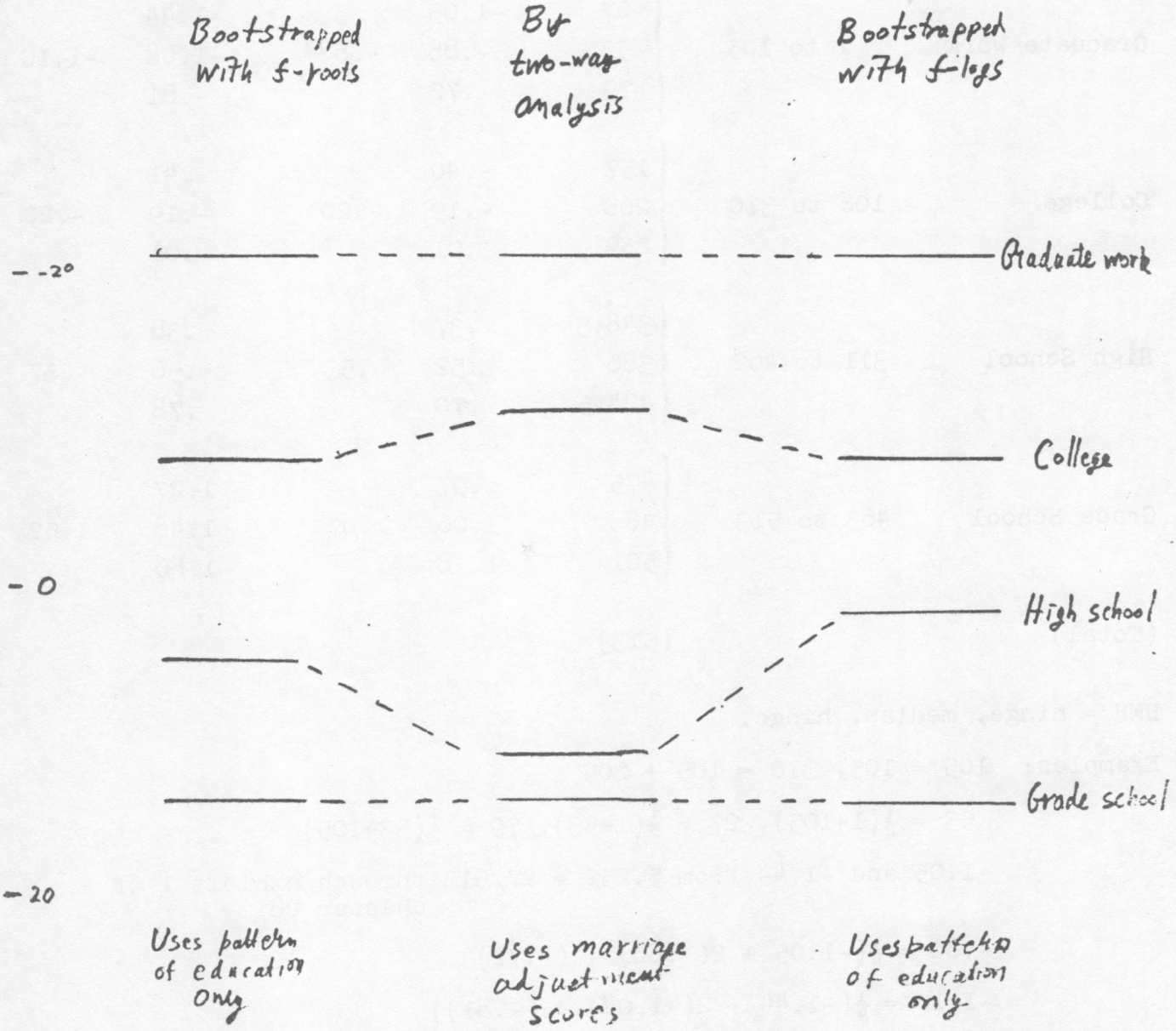
exhibit 26 of chapter 29 -- marriage adjustment

Comparison of scale spacings for education --

from marriage adjustment scores and from

distribution of education only.

(End scores forced to match, see exhibit 25)



Bootstrapping will not always work well, but it works well surprisingly often. We ought to feel free to:

- try it whenever we have nothing better.
- use it as a rough comparison when we think we have something better -- even much better.

****

29H. How far have we come?

This chapter has been devoted to making ordered responses more useful. If an ordered response can only be "higher" or "lower", we are back to a counted fraction and the techniques of the last chapter. When, as is even more often the case, we have more than two alternatives, we have a choice of cutting points. It usually pays us to find an appropriate way to make use of more than one such.

One simple approach is to find an f-log or f-root corresponding to each cutting point. We can then combine our carefully expressed responses across cutting points, something that amounts to analyzing a two-way table with

cutting point

and

circumstances

as factors. Sometimes this table is complete enough to let us be sure that our choice of expression gives us "parallel motion" -- lack of interaction, that is. At others, we can only say that we have been able to paste together a long scale.

If we want to display and/or compare counts of different ordered response, we can again use split figures.

When we deal with many accounts of quantitative responses, we often do well to begin by treating our responses as only ordered, and then to find a new expression for the cutting values that fits in well with the data. This happens most often when the quantitative response is coarse-grained.

When we come to look hard at two-way tables, we are not surprised to find a large number of them really have

cutting point for some response

and

circumstances

as the two factors. Getting rid of the cutting-point factor by finding f-logs or f-roots and analyzing the resulting two-way table to provide one or both of

combined scores for circumstances

and

a scale for the cutting points,

is then almost always the thing to do.

When the circumstances factor is also ordered, we can take advantage of this fact in our analyses.

****

29K. <u>More pictures for the rain forest (optional)</u>

We have not taken our pictures of the rain forest data as far as we might. We can use a reasonable variety of approaches. Exhibit 27 shows the fit alone in terms of seven half-cigars sliding up and down across the cuts. Such a picture is reasonably useful, but has its limitations. If it is going to look reasonably well, the half-cigars have to be about equally spaced.

We can always go one step further with depicting two-way analyses where one factor conceals a response. One reason we gave up plotting data and fit together on the same graph was the stubborn way in which the paths we wanted to draw would cross. This cannot happen when the versions of one factor are alternative cuts in an ordered response.

So we can, as exhibit 28 shows for the rain-forest data, always display both fit and data on the same picture. An alternate approach is to:

- condense most half-cigars to vertical lines.
- keep these level and set the fit tilt.
- space them so the fit will be straight lines.
- eliminate all unnecessary lines or parts of lines.

Exhibit 29 shows the result for the rain-forest data.

exhibit 27 of chapter 29 -- rain forest

Fit only, as shown by f-root split cigars, for the

diameter counts from Moraballi Creek

(see exhibits 20 to 21)

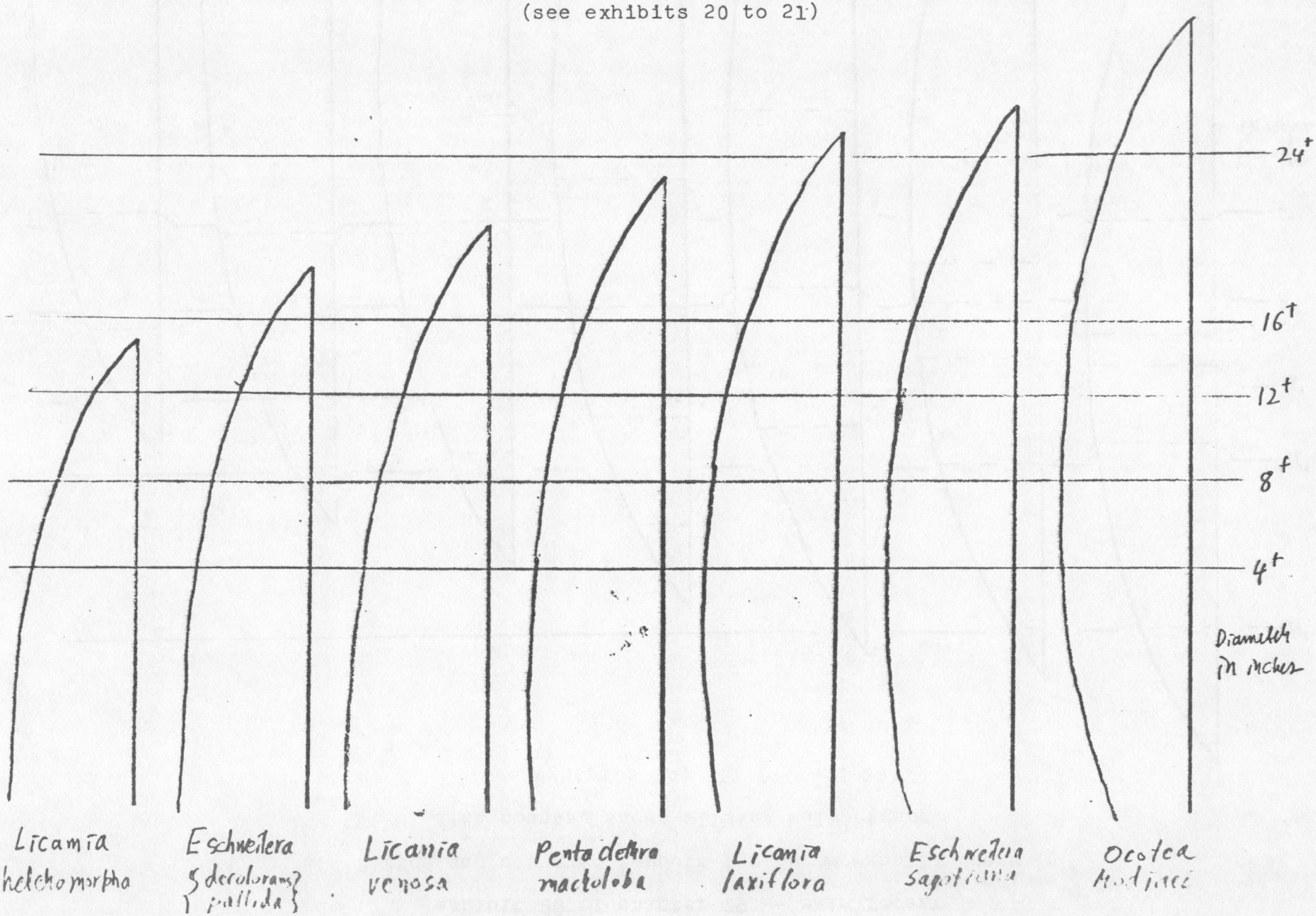


exhibit 27 of chapter 29 -- rain forest

exhibit 28 of chapter 29 -- rain forest

Fitted and observed f-roots for the Moraballi Creek

size counts, shown sliding half cigars

(see exhibits 20, 21, 27, and 29)

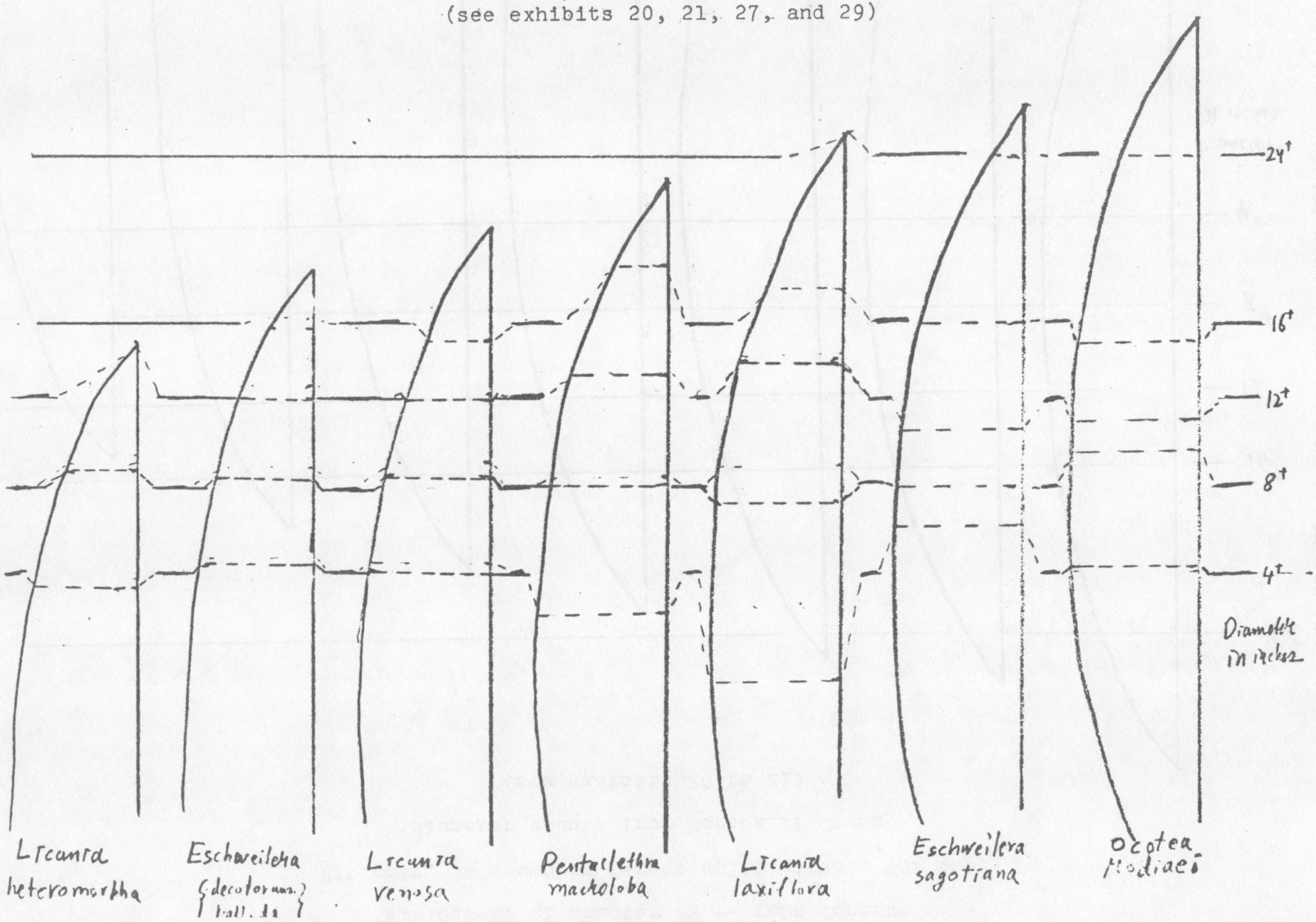
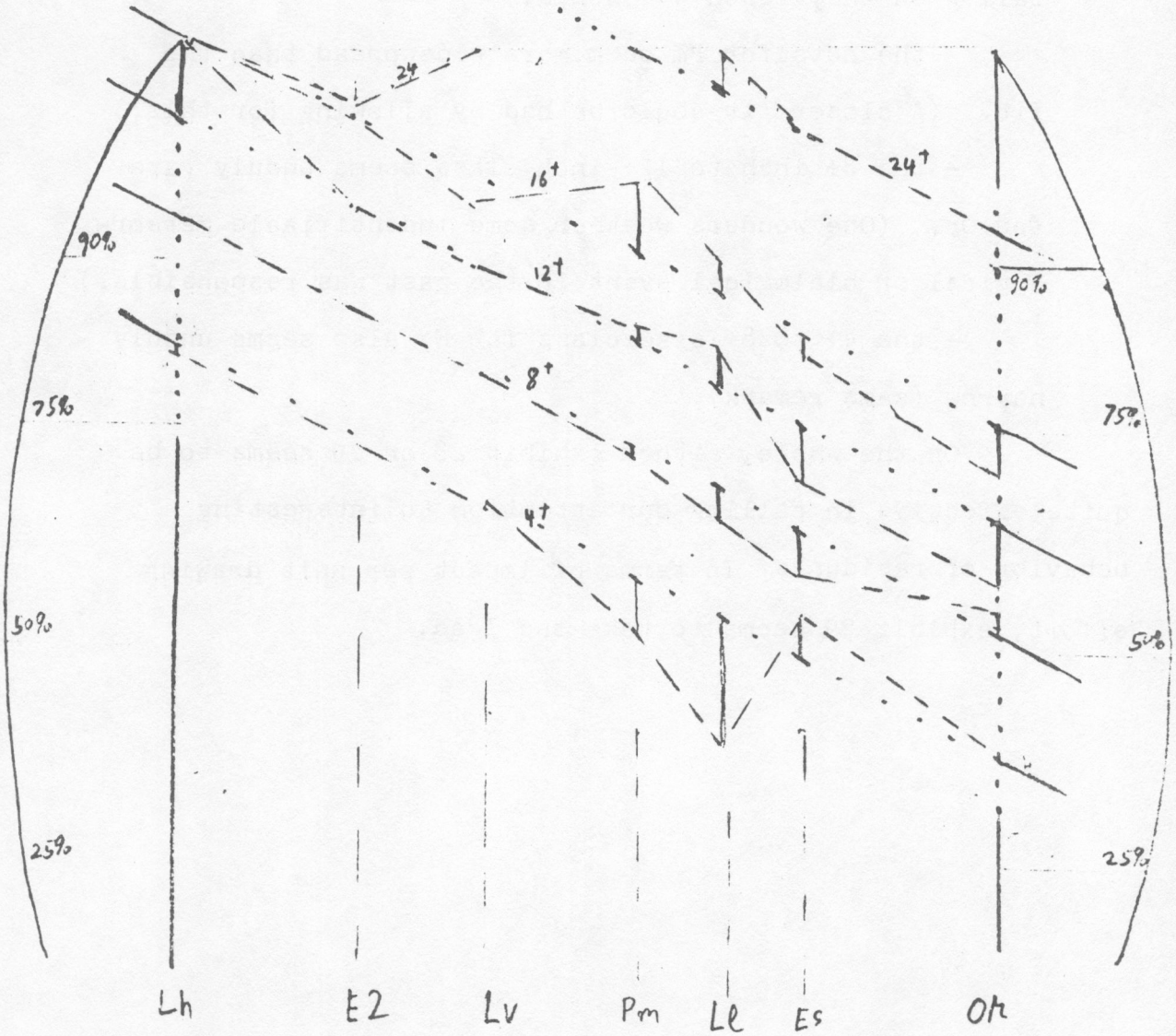


exhibit 28 of chapter 29 -- rain forest

exhibit 29 of chapter 29 -- rain forest

Observed and fitted f-roots in a squeezed half-cigar plot designed to make fits straight lines

(see exhibits 20, 21, 27 and 28)



Dashed lines are observed.
Slants at ends of short verticals are fitted.
Horizontal spacing determined by fit.

If we look only moderately hard at either exhibit 28 or exhibit 29, we see some things that we could have seen in exhibit 20, had we looked hard enough, namely:

- the 4+ inch cut for Ll falls strikingly low in its half-cigar. (Are Ll trees frequently under 15 feet tall when they reach 4+ inches?)
- the cuts for Pm seem more widespread than the fit. (A closer fit could be had by allowing for this.)
- the 8+ inch to 12+ inch class seems unduly rare for Or. (One wonders whether some indentifiable meteorlogical or biological event in the past was responsible.)
- the 4+ to 8+ size class for Es also seems unduly narrow (same remark).

On the whole, either exhibit 28 or 29 seems to be quite effective in calling our attention to interesting behavior of residuals. In terms of impact per unit drawing effort, exhibit 29 seems to take the lead.

****

29P. Problems for chapter 29

1. Do problem 1 of exhibit 5.

2. Do problem 2 of exhibit 5.

3. Do problem 3 of exhibit 5.

4. Do problem 4 of exhibit 5.

5. Do problem 1 of exhibit 6.

6. Do problem 2 of exhibit 6.

7. Do the problem of exhibit 9.

8. Do the problem of exhibit 22.

9. Exhibit 30 shows the after-career records of 223 delinquent boys divided according to their intelligence as measured by I.Q. tests. Analyze the table, comment.

10. Analyze one or more of the other tables given by David and Pearson at pages 64 to 66.

11. In 1954, Cochran gave the data in exhibit 31 concerning clinical response to the treatment of leprosy. Analyze. Comment on results.

12. In 1906, de Helguero gave the data in exhibit 32 concerning the number of flowers per inflorescence of Cichorium intybus. Analyze. Comment on results.

13. In 1907, Pearson quoted from Yule the data in exhibit 33 on naming tints. Analyze. Comment on results.

14. Find a set of data that interests you and for which the analysis of this chapter is appropriate. Make the analysis or analyses. Plot the results. Discuss what you see.

exhibit 30 of chapter 29

After-career Records of Delinquent Boys

Intelligence	Doing well	Fairly well	Badly	(T)
Normal and above	23	5	12	(40)
Dull normal	20	13	10	(43)
Border line	22	20	17	(59)
Mentally deficient	26	26	29	(81)
(T)	(91)	(64)	(68)	((223))

(T) = totals.

Source: F. N. David and E. S. Pearson 1961. Elementary Statistical Exercises. London; Cambridge University Press. (Esp. Table 63 on page 64.)

exhibit 31 of chapter 29

Treatment of leprosy by sulfones and streptomycin
(Entries are counts)

clinical change*	Initial degree of infiltration	
	0 to 7	8 to 15
Marked	11	7
Moderate	27	15
Slight	42	16
Stationary	53	13
Worse	11	1

*Improvement, if not stationary or worse.

Source: W. G. Cochran 1954. Some methods for strengthening the common chi-square tests. Biometrics 10: 417-451. Table 6 on page 435.

His source: Leonard Wood Memorial of the American Leprosy Foundation.

exhibit 32 of chapter 29

Number of flowers per inflorescence

for Cichonum intybus

(Entries are numbers of inflorescences)

per inflor.	# of inflorescences on plant: one	two	≥ three	(sum)
8	2	1	1	(4)
9	7	5	9	(21)
10	28	13	14	(55)
11	107	80	72	(259)
12	126	98	86	(310)
13	84	76	73	(233)
14	27	16	47	(90)
15	7	7	8	(22)
16	1	3	1	(5)
17	-	1	-	(1)
(sum)	(389)	(300)	(311)	((1000))

exhibit 33 of chapter 29

Comparison of names assigned to tints
(Two tints on each card)

Name given to lower tint	Name given to upper tint: Light	Medium	Dark	(Sum)
Light	850	571	580	(2001)
Medium	618	593	455	(1666)
Dark	540	463	457	(1453)
(Sum)	(2008)	(1620)	(1492)	((5120))

Source: K. Pearson 1907. Reply to certain criticisms of Mr. G. U. Yule. Biometrika 5: 470-476

His source: G. U. Yule 190?. On the influence of bias and personal equation in statistics of ill-defined qualities: An experimental study. Abstract in Proc. Roy. Soc. (London) Ser. A 71: 337-339. Table on page 338. (Full account in J. Anthropological Inst. 36: ???).